

全国最低交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易所手续费
任何地方费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进：<http://www.7help.net>



Mc
Graw
Hill
Education

金融知识堂

打开高频交易 ALL ABOUT HIGH-FREQUENCY TRADING 的黑箱

[美] 迈克尔·德宾 (Michael Durbin) 著 谈效俊 赵鲲 朱星星 译
南华期货研究所

肯·格里芬旗下城堡集团前高频交易系统开发者
从实操层面全面解构交易世界的王者：高频交易



机械工业出版社
China Machine Press



打开高频交易 ALL ABOUT HIGH-FREQUENCY TRADING 的 黑箱

[美] 迈克尔·德宾 (Michael Durbin) 著 谈效俊 赵鲲 朱星星 译
南华期货研究所



机械工业出版社
China Machine Press

图书在版编目（CIP）数据

打开高频交易的黑箱 /（美）德宾（Durbin, M.）著；谈效俊等译. —北京：机械工业出版社，2013.11
（金融知识堂）

书名原文：All about High-frequency Trading: the Easy Way to Get Started

ISBN 978-7-111-44658-3

I. 打… II. ①德… ②谈… III. 证券交易 IV. F830.91

中国版本图书馆 CIP 数据核字（2013）第 259441 号

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问 北京市展达律师事务所

本书版权登记号：图字：01-2013-0597

Michael Durbin. All about High-frequency Trading: the Easy Way to Get Started.

ISBN 0-07-174344-8

Copyright © 2010 by The McGraw-Hill Companies, Inc.

This authorized Chinese translation edition is jointly published by McGraw-Hill Education (Asia) and China Machine Press. This edition is authorized for sale in the People's Republic of China only, excluding Hong Kong, Macao SAR and Taiwan.

Copyright © 2013 by The McGraw-Hill Asia Holdings (Singapore) PTE.LTD and China Machine Press.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including without limitation photocopying, recording, taping, or any database, information or retrieval system, without the prior written permission of the publisher.

All rights reserved.

本书中文简体字翻译版由机械工业出版社和麦格劳－希尔（亚洲）教育出版公司合作出版。

版权 © 2013 由麦格劳－希尔（亚洲）教育出版公司与机械工业出版社所有。

此版本经授权仅限在中华人民共和国境内（不包括中国香港、澳门特别行政区及中国台湾地区）销售。未经出版人事先书面许可，对本出版物的任何部分不得以任何方式或途径复制或传播，包括但不限于复印、录制、录音，或通过任何数据库、信息或可检索的系统。

本书封底贴有 McGraw-Hill Education 公司防伪标签，无标签者不得销售。

机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码 100037）

责任编辑：左 萌 版式设计：刘永青

三河市杨庄长鸣印刷装订厂印刷

2014 年 1 月第 1 版第 1 次印刷

170mm × 242mm • 13.5 印张

标准书号：ISBN 978-7-111-44658-3

定 价：40.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：（010）68995261 88361066

投稿热线：（010）88379007

购书热线：（010）68326294 88379649 68995259

读者信箱：hzjg@hzbook.com

前言

All About High-Frequency Trading

顾客：这些东西多少钱？

店主：一块五。

顾客：我买一点。

店主：那一块五毛一。

顾客：啊，你说了是一块五了。

店主：那时候我不知道你想买。

顾客：你不能这样做。

店主：这是我的地盘。

顾客：可是我需要买一百个！

店主：一百个？那就一块五毛二。

顾客：你也太坑人了。

店主：供求关系啊，同志。你买还是不买？

什么是高频交易？问得好！现在是时候给出一个回答了，因为好像每个人都在谈论这个话题，并且大家都爱憎分明。当你看到这种事情的时候，常常应该回过头来再问一下，这件事到底是什么？是不是只有股票市场有高频交易？还是股票衍生品（比如期权）市

IV

场也有？高频交易是不是包括某些自动化交易，让计算机代替人进行决策？高频交易是不是那些尖端交易公司的专利，它们会不会像上面的店主一样，因为想要更多的利润而随心所欲地操纵市场？好吧，既然没有人能确切地回答这些问题，我们还是给出自己的定义并且接着讨论下去。

一般而言，在买卖证券时使用高频交易，速度是成败的决定性因素，千分之几秒，或者几毫秒[Ⓐ]的差异，就可能成为盈利和亏损的分水岭。不仅股票市场有高频交易，股票期权和期货市场中也有。通常，并不是所有的交易一定都需要快速执行。比如，你看好一家公司未来几年的表现而买入其股票，或者你取出 401 (k)[Ⓑ] 账户中的钱买一部你 16 岁时就十分心仪的哈雷摩托车。但确实有很多交易策略取决于你能够多快地识别市场中的盈利机会，以及你能多快抢在其他人之前递交委托单。后文我们将介绍一些这类策略。

高频交易者是从做市商，或者叫专家的队伍中发展出来的。做市商主要的利润来源是他们报出的买卖价差。然而，高频交易者与做市商还有很多不同，这主要是因为随着十进制化的发展[Ⓒ]和技术的进步，高频交易者能够赚到的价差已经薄如蝉翼，甚至到了一分钱或者小于一分钱的程度。因此，高频交易者不得不进行大量的交易。实际上，较大的高频交易公司在市场中遨游，大口吞噬着各种微小的交易机会，就像巨鲸吞噬小虾一样。

Ⓐ 或许到你读到这本书的时候，微秒，即百万分之一秒，已经逐渐成为至关重要的时间间隔了。而纳秒，或者十亿分之一秒也只是时间早晚的事。

Ⓑ “401 计划”是美国从 20 世纪七八十年代开始发展的一种社会保障和养老金体系。

Ⓒ 十进制化指的是在 2000 年早期，交易股票从以美元的分数（例如，八分之一）进行计价转变成按分计价，这极大地降低了买入和卖出股票可能的价格差。

我们不难发现市场中存在高频交易的迹象，并且从事高频交易的公司也越来越多。图 0-1 给出了美国证监会（SEC）的一些计算结果，该图显示，从 2005 年到 2009 年，纽约证券交易所上市的股票的日交易量^①增加了近 3 倍，与此同时，每日执行的交易笔数增加了近 8 倍。两个数字合起来显示平均每笔交易的规模下降了，这也与图中显示的结果一致。^②如果市场中越来越多的公司在进行竞争性的做市，也就不难想象会出现这样的数据。^③然而，真正显示计算机开始占据主导地位的是图 0-2，该图表明，2005 年平均执行一笔交易所需要的时间是 10 秒钟，4 年之后，这个数字缩减到 1 秒以下。人是可以很快，但也做不到这么快。

高频交易要求的速度是人远远不能达到的。因此，高频交易必然是一种自动化交易。在高频交易中，计算机会实时地做出一些以前只有人类做的战术决策。从这方面来说，它与飞机的自动驾驶很类似。在现代飞机上，计算机即时地做出各种决策（比如，襟翼位置、空气速度，等等），以保持飞机飞行平稳和轨迹正常。计算机是基于人类数十年的手工飞行经验进行设计和编程的。人类摸索出来的驾驶飞机的策略和步骤用计算机软件的形式电子化表达出来。然而，驾驶室里面仍然保留有飞行员，负责监控自动驾驶仪，在安全的情况下打开仪器，对控制进行正确的设置，并且在必要的情况下代替仪器进行操作。

① 交易量指的是某给定时间段之内成交的股数。

② 美国证监会，“股票市场结构概念公告”（*Concept Release on Equity Market Structure*），1/14/2010。

③ 这种现象也与我们后文中要讨论的一种担忧相一致，即很多新增的交易量是高频交易公司彼此之间交易出来的，他们只是为交易而交易。

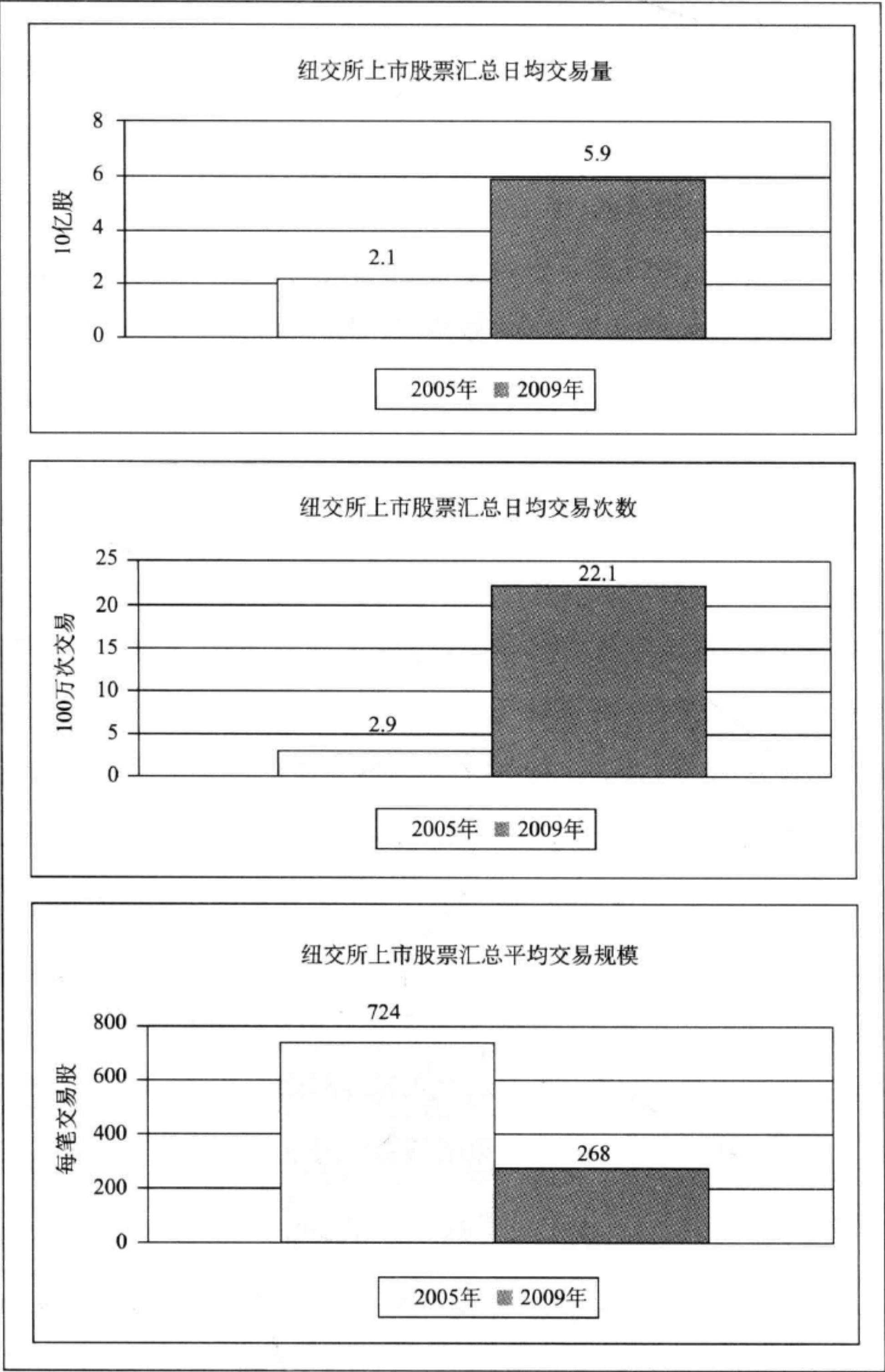


图 0-1 纽约证券交易所交易统计

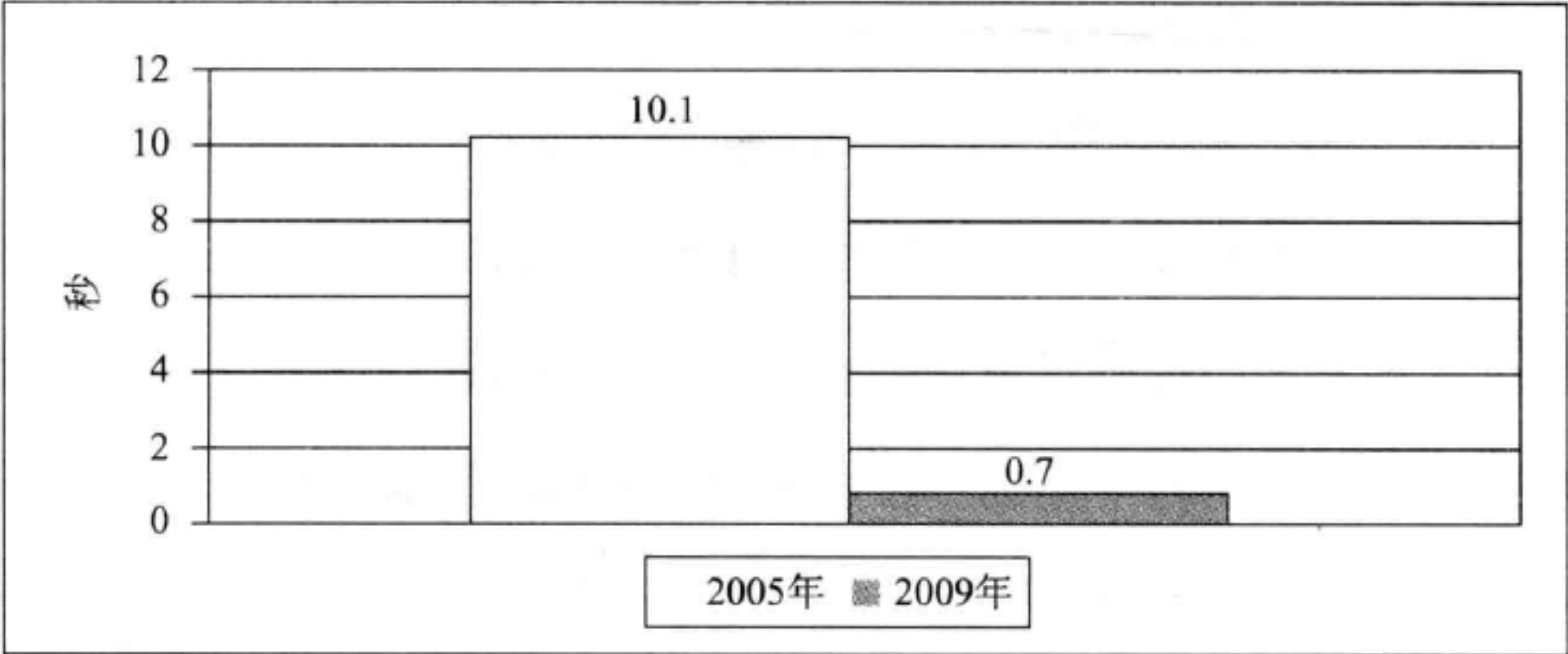


图 0-2 纽约证券交易所平均执行速度

注：针对小的、可立即执行的委托单。

高频交易有着相同的发展过程。发展这个词在此是恰到好处的，就像飞机的自动驾驶一样，高频交易也不是一夜之间发明出来的。高频交易代表了市场中技术因素的最新发展。这种发展已经持续了数十年并将一直持续下去。今天我们看来非常快的速度可能在将来看起来慢得可怜。

在这种广泛的意义下，很多交易都可以归为高频交易的范畴，但是有一类高频交易却在当下最为吸引眼球，然而它的名声并不见佳，也引发了人们对市场安全性和公平性的担忧。本书主要讨论的正是这类富有争议的高频交易，这类交易利用惊人的计算速度和网络速度去执行一种相当不入流的交易策略，或许可以称之为黄牛策略（scalping）。与球赛的黄牛贩子类似，这些证券黄牛们试图以一个价格买入，再迅速的转手以更高的价格卖出，或者在相反的方向上如法炮制，以赚取二者之间的所谓买卖价差。往往其中的一个价格来自错误定价的证券，例如，来自一只价格过高或者过低的股票。那么，高频交易者的主要工作就是发现这些机会并且在其他人之前抓

住这个机会（后文将讲述这是怎么做的）。

本书是对高频交易领域的一个概述，在书写和组织上，本书针对的是那些对交易没有任何先验经验的读者。首先，我们将对证券市场上各种类型的证券以及它们之间的主要关系进行通览，以便为读者理解高频交易打下基础。我们还要花一点时间来弄明白在交易所实际发生了什么事情，并且熟悉和了解交易实际发生的地方——委托单。在浏览了交易的基本概念之后，我们将交易者分成四种类型，我们称之为投资者、做市商、套利者和预测者，接下来我们将讨论他们各自的策略，并且在这个大背景下观察高频交易者所处的位置。前文已经提到，现在公众对高频交易的风险不乏关注，我们也将对此忠实地阐述，并且尽量对高频交易可能的好处和一些高频交易公司传说中的可疑行为平均着墨。

这本书里面的信息和论断有些来自于我的个人经验，有些来自于交易者、交易所工作人员，以及行业中与我交谈的好心人。据我所知，本书的所有内容都不涉及任何可以称之为某家公司私有财产的东西。虽然高频交易者都极尽所能保守秘密，但这个行业还是有很多公共知识，如果你对金融世界这个角落还很陌生的话，这些公共知识足够你学习一阵子的。噢，在我忘记之前，我还得负责任地说，直到本书出版，我们也没有办法确认本书中讨论的策略哪些是合法的，哪些是非法的。目前，高频交易的适当性（无论是法律上还是道德上）正是证监会、华尔街，还有世界各地交易所一个辩论的焦点，所以我也不想在这上面说些什么，以免像是给出投资建议似的。不要把这里的方法想的太好，也请不要仅仅是因为在这里看到就采

用本书中所讨论的交易方法。

可能你对高频交易的赚钱能力有所耳闻。因此，我知道你们中间起码有一部分人会想：“我能不能在家里搞一套？”但愿你能吧。当然，你可以从百思买选购一台处理能力与高频交易公司服务器不相上下的游戏机。然而，你同样也可以从 **Sport Authority** 上买到和超级碗比赛用的一样的足球，在 **Guitar Center** 上买到同埃里克·克拉普顿（**Eric Clapton**）手上一样的吉他。当你读了这本书满怀雄心想大干一场的时候，别忘了高速计算机就像皮球和吉他一样，并不是这个游戏中需要的唯一的東西。

目录

All About High-Frequency Trading

前 言

第 1 章 搞砸了 / 1

第 2 章 交易 101 / 9

权益性证券 / 9

交易所 / 13

委托单簿 / 18

交易 / 24

交易者 / 27

第 3 章 交易策略 / 41

投资者与做市商 / 43

投资者策略 / 46

做市商策略 / 53

再谈投资者策略 / 67

套利策略 / 68

预测者策略 / 78

高频交易者的起源 / 86

第4章 提速 / 89

购买还是自建 / 92

让设计适应变化 / 95

分布式与负载平衡 / 98

设计最高效的软件 / 100

出色的通信机制 / 104

内核之外的任务处理 / 107

并行处理 / 111

直连交易所接收数据和执行指令 / 112

原生 API / 114

共置 / 116

高速数据通道 / 118

实时监控 / 121

第5章 揭开面纱 / 125

高频交易服务器的位置 / 126

基础设施：系统通信和数据库 / 129

各种组件 / 130

交互场景 / 156

仪表盘 / 163

合上面纱 / 163

XII

第 6 章 有关高频交易的争论 / 165

好处 / 165

担忧 / 168

玩过界了？ / 177

未来可期？ / 183

现在如何？ / 185

术语表 / 187

第 1 章

All About High-Frequency Trading

搞 砸 了

一直不知道这事什么时候能为人所知，至少我在 2003 年的时候是了解了一些。那时我去了城堡投资集团（Citadel Investment Group）工作，这是肯·格里芬（Ken Griffin）在芝加哥的一家神秘对冲基金，我的工作任务是帮助他们建造一个在美国股票期权市场运行的高频交易系统。当时美国只有一个完全自动化的期权交易所，即位于纽约的国际证券交易所（International Securities Exchange, ISE）。在 ISE 于 2000 年开业之前，没有哪家期权交易所允许做市商（一类我们马上就会讲到的交易商）通过电子系统报价——申买买入和申卖卖出，这些对做市商而言是十分重要的报价。在大多数地方，各个交易所仍然是由交易员穿着耐用的鞋子，整日站在公开喊价交易池里，通过大声叫喊和挥动胳膊进行口头报价。ISE 的创建者大卫·克莱尔（David Krell）和加里·卡茨（Gary Katz）看到了变革的机会，并且他们的成功也表明他们做出了正确的选择。ISE 是典型的游戏规则改变者。它突如其来的成功迫使其他传统的、依靠交易大厅的期权交易所也开始制定发展电子报价系统的计划。而肯·格

2 打开高频交易的黑箱
All About High-Frequency Trading

里芬希望城堡基金能够抓住这一切。

此时我已经管理金融系统开发（主要是给证券衍生品定价）工作好几年了。然而，城堡基金的系统不仅需要同时计算成千上万只期权的价格——这堪称它们的一项独门绝技，它们还需要实实在在地以超人的速度向市场注入申买和申卖委托单流。定制的报价引擎每天都不知疲倦地向市场发送数以百万计的报价，每份报价都约定以某个指定价格买入或者卖出期权合约，买卖的价格也是计算机软件算出来的。在报价机器辛勤工作的同时，“电子眼”实时地扫描着其他市场参与者每天数以亿计的报价和委托单。这就好比站在消防水龙管跟前，在每滴水落地之前都把它检查一遍。当电子眼（或者叫EE）发现有人以高于期权价值的价格买入或者以低于期权价值的价格卖出，它将在相反的方向提交一张委托单并赚到一点点利润。

当你目睹机器把市场的无效之处清理得一干二净，这可真是一种华丽的体验。我当时很想跟家人和朋友们说说，告诉他们迪尔伯恩大街和亚当大街交叉口的这幢闪闪发光的玻璃大厦里面到底发生了什么。但是像城堡基金这样的雇主和员工签订的保密协议往往非常、非常的起作用。在这一行，所有人都知道，今天嘴巴没管牢，明天收到粉纸条[Ⓐ]。因此，我还是决定像其他人一样，把嘴巴关紧，只和我在37楼工作的几个同事讲话。[Ⓑ]

2005年我离开城堡基金的时候，其期权做市系统的交易量已经超过全美期权交易总量的10%，这意味着它们每天交易超过100万份期权合约。这个做市系统是一个不大的团队开发的，开发团队的

Ⓐ 粉纸条是字面翻译，引申为解雇通知书。——译者注

Ⓑ 即使是在城堡集团内部，也是该你知道的要知道，不该你知道的不要知道。有一次我想搞明白我们的一个重要的波动率模型中参数的意思，我问一位量化分析师：“‘v’是什么意思？”“意思是‘v’。”她答道。“呃，那‘h’呢？”“意思是‘h’。”我一连问了好几个字母，得到的都是这样半开玩笑式的回答。

人数比芝加哥小熊队先发阵容多不了多少。在此后的3年之内，有报道说它们已经占据了市场份额的30%，在市场中处于一种支配性的地位。美国的期权市场已经被少数几家类似于城堡基金的神秘公司的超级机器所控制，交易已经不再与交易员有关，然而，似乎外界对此一无所知，或者是漠不关心。

到了2009年，事情有了变化。人们从2008年市场崩溃的余波中爬起，擦干身上的血迹，思考着金钱到底去向何方，有传言说有大约200亿美元到了一帮叫做“高频交易者”的腰包里。像城堡这样的公司对“高频交易者”这个词已经司空见惯了，但外界对这个词还很陌生。现在，仅是有人能在2008年大发横财就已经足够令人发指，如果再深想一步，恐怕不少人要去厨房拿菜刀了：高频交易是否诡谲无行，它会不会引发市场崩溃，或者是在崩溃时推波助澜？毕竟，很多人都认为自动化交易（后来称为程序化交易）是导致1987年市场崩溃的元凶。计算机是不是又干了同样的事情？

有传言说股票市场50%的交易量都是高频交易的计算机彼此交易出来的，还有人说是60%甚至70%，这么多交易量可能是为了从每笔交易中得到一点回扣，也就是人们常说的返还。高频交易公司不会长时间持有股票。一旦买入，它们会立即卖出，有时一天之内同一只股票会来回买卖成百上千次。这到底算个什么？这些交易都是以堪称诡异的速度完成的，据说如果不是因为这些交易，市场价格波动根本就不会有这么大。这么做能有什么好的结果？一名高盛前雇员因为被控偷窃公司高频交易代码而被捕，这让大众的神经更加紧绷。这家投行宣告了一条不祥的消息：“存在一定的风险（知道如何使用这个程序的人拿它以不公平的方式操纵市场）。”^① 用计算机

① “高盛可能因前员工偷窃代码损失上百万美元”，Bloomberg，2009年7月7日。

4 打开高频交易的黑箱

All About High-Frequency Trading

代码操纵市场？我的天哪！这到底是怎么回事？

关注高频交易掌故的人会从现代交易词典中学到很多新名词，遗憾的是没有哪个词会给人留下好印象。高频交易公司可能使用了一种叫做**闪现委托单**（flash order）的东西提前看到客户的委托单，并且用偷看来的信息为自己的交易牟利。这难道不是**抢先交易**（front-running）吗？难道不是违法的吗？这些闪现委托单劫匪们只有 30 毫秒来进行这些下流勾当，不过他们可是花了大价钱把交易服务器共置（colocate）在交易所计算机所在的数据中心里面的，30 毫秒对他们来说已经绰绰有余了。他们还能由此抢在所有投资者之前把委托单递交到交易所。玩法本身就不公平，有人不同意吗？

高频交易的精细鬼们并不会止步于偷看闪现委托单这种事情，据说他们还提交一种叫做**不成即撤委托单**（immediate-or-cancel order, IOC），他们提交这种委托单并不是真正为了交易，而是用于探测其他投资者想要交易的价位，高频交易者可以利用这些信息操纵市场价格往不利于投资者的方向移动。哇！这些高频交易公司到底有没有权利离交易所这么近并且以光速进行交易？没人知道，甚至市场中识别出高频交易者都几乎是一件不可能的事。他们并不需要自己投资建设高频交易的基础设施，只需要借助经纪商与交易所之间的连接，以**直接市场接入**（direct market access, DMA）或者**裸接入**（naked access）的方式匿名进入市场，就像披上了一件斗篷，从此可以在市场中以光的速度为所欲为^①。

“老兄”，你几乎可以听到人们疲惫不堪而又气急败坏的吼叫：“我们的股票市场怎么了？”市场不再是以前那个样子了吗？不再是

① SEC 主席玛丽·夏皮洛把 DMA 比作把没有车牌的法拉利车子的钥匙借给可能没有驾照的人。

简简单单地通过买入一只股票并且持有一段时间来投资一家公司了吗？我们还抱有这种想象是不是有些太傻了？或许我们已经成了那些拥有高速计算机和出色数学技巧的人们的玩物，就像一个街上徘徊的傻子，任由小摊小贩们打量和哄骗。市场还有正常的功能吗？难道它已被冷酷的信息技术和失控的贪婪所绑架？

看起来确实是这样。

2008 年对高频交易者来说是不折不扣的奇迹之年，对于高频期权交易者更是如此，这个消息会让很多人感到震惊，甚至怀疑是不是听错了。一个猎头告诉我，他的一个客户，也是一家高频期权做市公司，在 2008 年挣到了超过 8 亿美元。另外一家有名的公司据说结清的净利润是 13 亿美元——这仅仅是交易期权的利润。谁知道他们在高频交易那块做了什么手脚？

作为一个曾经的业内人士，我做过一些枯燥的 P&L^① 计算，发现这些传说中的数字是完全可信的。2008 年期权的交易量是 30 亿手^②，假设每交易一手，做市商的利润是 2 美元^③，那么整个市场的利润大约是 60 亿美元。即使这个估计偏高，市场的总体利润也有 50 亿或者 40 亿美元。鉴于分这块蛋糕的做市商不多，完全可能有人分到的份额会达到 10 亿美元。不过，仅仅是大部分人在 2008 年损失惨重，而有人能赚得盆满钵满，并不能说明这些人所取的都是不义之财。当技术的进步使行业发生变革的时候，不论是个人电脑，还是汽车

① 即盈亏 (profit and loss)。

② optionsclearing.com/webapps/historical-volume-query。

③ 此处假设平均每次交易有 0.03 美元的优势（比如，做市商以 5.03 美元的价格卖出一份理论价格为 5.00 美元的期权）。设其中的 1/3 被做市商用作对冲费用（为保持 delta 中性调整股票持仓、波动率估计的偏差，等等）。大多数期权都对应 100 份标的证券，因此实际成交价格时期权报价再乘以 100（对于标价为 5.03 美元的期权，做市商卖出收到的金额是 503.00 美元，也就是说交易利润是 3.00 美元，扣除费用之后还剩下 2.00 美元）。

制造，或是石油精炼，都有一小部分先行者（盖茨、福特、洛克菲勒）从他们的投资中赚取不可计数的财富。

接着汽车这个比方，来说说安全性的问题。在 20 世纪初，不用马拉的机动车辆刚刚出现，最高时速仅能达到每小时 15 ~ 25 公里，并且路上这种新生事物也为数寥寥。然而，随着技术、需求与自由市场企业的融合，汽车的最高时速与日俱增，马路也开始拥堵起来。车毁人亡的事故屡次发生，这种新型装置的杀伤力凸显了出来。政府和行业开始关注设计安全性问题和汽车管理问题。人们制定了最高速度限制，发明了安全带，制定了相关法律，警察开始开出罚单。完全可以说，高频交易也需要进行类似的反思，以保护人们免受新技术的伤害。

汽车革命和高频交易之间还有一个显而易见的相似之处。在汽车发展的早期，有些人并没有把汽车看作一种便利工具，相反，把它看成是助长犯罪的工具。如果约翰·迪林杰（John Dillinger）的逃跑装备是一辆手推车的话，他还能这么逍遥法外吗？银行劫匪犯罪之后乘坐汽车逃之夭夭，要是没有汽车，他们也就不能进行这些犯罪活动了。相应地，警察也配备越来越好的汽车，法律也进行了扩展和修订。人们是不是也能利用高频交易的工具去做一些他们之前（比如，10 年之前）做不到的事情然后溜之大吉呢？这并非天方夜谭。是否有必要重新考虑一下证券市场监管和法律实施的问题了？或许现在正当其时。

必须指出的是，构建高频交易系统的高频部分可不像逛商场那么容易。建造一个实际可用的高频交易系统非常复杂，其中要做的事情不止一件，要掌握的工具也不止一种，就像建造一艘远洋客轮一样，远远不是一锤定音那么简单。从雇用适当数量的交易、数学、

软件开发等领域的奇才开始，在追求一缕缕微小利润的道路上，高频交易者不会放过任何一个细节。他们不使用第三方开发的授权产品，而是自己编写软件；他们根据自身需求制作各种个性化系统，并为此投入大量的人力财力。噢，还有，构建和维护这些错综复杂的巨无霸系统花费不菲，即使是沃伦·巴菲特看了都得皱一下眉头。

高频交易公司在软件开发的过程中大量应用了软件工程方面的知识，这使它们得以开发出功能容易更改，并且代码极其高效的软件。它们的软件在设计上必须保持容易更改的特性，因为只有这样才能跟得上那些同样装备精良且雄心勃勃的竞争对手们。它们的系统会借用复杂事件处理领域的模式，是所谓分布式，实时系统的完美典范，系统中数以千计的程序有条不紊地运行在适量数据center里适量的计算机上。它们一反过去几十年来将大量进程扔给CPU（中央处理器）的传统，转而使用专门硬件去处理某些计算问题，这种做法看上去有些倒退。它们甚至在金融计算中用上了图形处理器（本质上是一种游戏卡），这是因为图形处理器具有大规模并行计算的能力。

高频交易公司从来不想通过第三方接收市场数据（也就是那些举足轻重的分笔数据流）或递交交易委托单，它们嫌第三方的连接太耗时了。因为速度的原因，它们会坚持直接接入交易所。只要条件允许，它们就会弃用行业标准的FIX协议^①，而采用交易所原生的应用程序接口（API）与交易所进行通信。它们还对交易所计算机（或者叫撮合引擎）所在的位置了如指掌，并且会在相同的数据中心租

① FIX，即金融信息交换（financial information exchange），是一种用于电子化传递证券交易信息的行业标准协议。FIX本身是一个非常好的东西，它大大简化了递交交易委托单这种工作，但也有人认为它会在委托单报入的过程中增加一点十分微小的延时。

8

打开高频交易的黑箱

All About High-Frequency Trading

用机位以共置自己的服务器，所有这些仅是为了能够领先下一个参与者百万分之几秒而已。

费劲？烧钱？那是当然。但是想想吧，一台这样的赚钱机器每年能带来数以亿计的利润，也就难怪越来越多的公司会不顾一切地想造出一台了。

第 2 章

All About High-Frequency Trading

交易 101

让我们把重要的事情放在前面做：在讨论高频交易之前，先回顾一下传统的交易方式。后文我们会谈到美国市场上交易的权益性证券，以及证券交易所中所发生的各种事情——谁在做什么、为什么这么做，以及不同参与者之间有哪些天然的相互作用等。先回顾一下传统的交易方式对我们未来论及此类问题会大有帮助。

权益性证券

很多有关高频交易的讨论都与已上市的权益性证券（或者叫做股票）有关。在权益性期权（股票期权）和股指期货这两个与股票关系密切的市场上，高频交易同样十分活跃。美国的股票市场、期权市场和期货市场之间的联系十分紧密，它们共同形成了一个巨大的权益类证券市场，高频交易就在这个大“超市”里大展拳脚。事实上，高频交易出现在这些市场中并不是偶然的，有证据表明，正是由于高频交易，这些市场才能如此紧密地联系在一起。

本书将主要关注股票市场中的高频交易，有时因为读者可能想了解更多信息，我们也会涉及期权和期货市场，不过这些都不是理解高频交易所必需的。并且，在后文你会看到，股票市场中的一些交易策略实际上是期货和期权市场驱动的，因此，读者也有理由多少了解一点这些市场的情况。高频交易当然不会局限于美国的权益性市场。在非美国市场，以及非权益性市场，高频交易的影响力也越来越大，比如商品市场、利率市场、货币市场（亦称外汇市场，或者叫 FX）。虽然本书并不会直接涉及这些市场，但是其中讨论的概念对这些市场也是适用的。

股票

股票，当然代表的是对公司的一份所有权。一般而言，美国市场一天的交易量大约在 90 亿股左右^①。这些股票不仅包括大约 5 000 家上市公司——比如谷歌（代码 GOOG）、美国铝业（AA）、摩托罗拉（MOT），等等——还包括交易所交易基金，或称 ETF。ETF 与传统股票的交易方法类似，不过其价值是与某只指数挂钩的。最近几年 ETF 变得相当普及，有些 ETF 的交易量已经能够与最活跃的股票的交易量相媲美。

ETF 有时也称为**指数跟踪股票**（index tracking stock）。例如，按定义，Spider（代码为 SPDR）这只 ETF 的价值应该约为标普 500 股票指数的 1/10^②。另外还有一只跟踪纳斯达克 100 指数（QQQQ，又称“the Qs”）的 ETF 也很流行。ETF 与其他单个公司的股票列在一

① 该平均值根据 batstrading.com 提供的 2009 年 10 月 19 日至 2009 年 11 月 13 日各市场交易量数据求得。

② 按定义应该是严格的 1/10，但是由于交易 Spiders 与交易基础指数的费用差异等原因，Spider 的价值与理论价值常常会有几分钱的差异。

起，甚至也派发其基础股票所产生的累积股息。传统类型的 ETF 有好几十只，不过其中只有少数几只在市场中处于主导地位。还有一类新的 ETF 称为杠杆 ETF，这类产品会以一定的比例（通常是 2 倍或者 3 倍）放大传统 ETF 的收益。

期权

股票期权是一类衍生证券。在股票期权中，投资者通过支付一定的权利金，来换取未来某个时间以约定执行价格买入（*看涨期权*）或者卖出（*看跌期权*）某股票或者 ETF 的权利。股票期权的一个常见变种是指数期权，其标的物不再是某只股票或者 ETF，而是一个指数（你可以把它看成是一篮子股票），例如标普 500 指数。对于一只股票或者某个指数，可能会存在上百个不同执行价格、到期时间和类型（看涨或者看跌）期权合约。因此，期权合约的数目要多于，甚至是远远多于股票的数目。

这个现象也正是高频交易适用于期权交易的原因之一。期权合约的数量非常之多，相比于人类交易员而言，计算机在跟踪这些合约方面更有优势。高频交易与期权结合紧密还有一个原因，就是股票价格与其对应的期权价格（如果是指数期权，则对应的是指数点）之间有相对确定的关系。期权价格由多个因子决定，其中最为明显的就是标的股票的当前价格。当股票价格变化时，期权价格就会随之发生改变，期权价格可以说是股票价格衍生出来的（也因此得名 *衍生品*）。只有计算机才能胜任实时地根据股票价格来重新计算期权价格（有时计算机也跟不上股价的变动。后文我们将看到这个过程所产生的延迟带来了怎样的套利机会）。

在过去几年，股票期权和股指期权已经非常盛行，2009 年每日

交易量达到约 1 400 万张。^①大部分期权合约都赋予交易者交易 100 股基础股票的权利。因此，每日交易的期权所涉及的股票达到约 14 亿股。

期货

股指期货是另外一种可以约定未来某个时间以某个交割价格买入或卖出一篮子指数标的股票的衍生品。与期权类似，买卖股指期货时，你并不需要真正买入或者卖出股票篮子（股指期货通过现金交割），但你所实现的盈亏与真实买入或者卖出股票篮子类似，并且盈亏都是逐日结清的。因此，在这类期货上持有头寸实际上也就相当于对市场走向下了一个赌注。

股指期货中最活跃的是在芝加哥商品交易所（CME）交易的迷你标普 500 指数期货合约（E-mini S&P 500 futures contract, ES）。假如你在 1 月底以 1 100 美元的价格买入一份 3 月交割的 ES 合约，实际上相当于你承诺以 55 000 美元的价格买入标普 500 指数 500 只成分股。^②在购买合约时，你不需要支付任何费用。^③合约的交割价主要是由指数点数、利率、到期时间和股票的预期股息所决定的，交割价使得合约当前价值为零。在整个交易阶段，交割价会持续发生变化，你持有的 ES 合约价值也会相应的发生变化。如果指数上涨，你的持仓价值会上升；如果指数下跌，你的持仓价值会下降。

假设到了 3 月的合约交割日，指数达到 1 200 美元，那么你持有的 ES 合约的标的股票篮子现在的价值为 60 000 美元，也就是比你

① 期权清算公司（Option Clearing Corporation）有一个很好的网站（optionsclearing.com），在这个网站你很容易就可以查到美国期权市场交易了多少期权。

② 一手 ES 合约的名义价值是 50 美元乘以标普 500 指数的指数点。

③ 你确实还是需要支付经纪费用和其他一些交易费用，但是合约本身并没有任何基础价值。

当初承诺支付的价格高出 5 000 美元。假设你在交割之前一直持有合约，那么你就会获得大约 5 000 美元的利润（由于期权合约的逐日结算制度，你实际获得的会略少于这个数字。在这个例子当中，你在交割日之前至少会得到一部分盈利，这部分将是按照现行利率对收益的折现）。

ES 听起来和期权合约很像，但实际上期权与期货有一个重要区别。如果你当初是承诺卖出（例如，不是持有 ES 期货合约的多头，而是持有空头），那么你同样也会亏损大约 5 000 美元。与期权不同的是，期货的持有者有以约定价格买入或者卖出标的资产的义务，而不管到时现货价格是上涨还是下跌。还有一点需要记住的是，期货合约在起始的时候是没有基础价值的，而期权合约无论在什么时候总是具有一定的价值。期权的价格可以视作为保护资产价值损失而支付的保险金。

现在你该可以猜到为什么股票市场、期权市场和期货市场如此紧密关联了。想象一下标普 500 股票指数 500 只成分股中的一只，当这家公司股票的外在价值发生改变时，它当然会非常明显并且直接影响到你买卖这只股票的价格。不仅如此，它还会影响到这只股票所对应的期权合约的价格。并且，这只股票作为标普 500 指数的成分股，其价值改变也会影响到指数点位，进而影响到指数期货合约的价格。因此，股票、期权和期货的价格有天然的内在联系，而只有计算机的速度才能快速同步地进行这些数学计算。

交易所

交易所是传统的进行股票交易的场所。交易所是一个有形的地

14 打开高频交易的黑箱

All About High-Frequency Trading

点，参加交易的人在此聚集，买方和卖方彼此交易所谓的上市的股票。在交易所市场之外还有一个场外市场（OTC），在场外市场中，交易双方找到交易对手，并且进行私下谈判成交。^{⊖⊖} 这里我们将不涉及场外市场。

还有一种交易机制称为内部化（internalization），在这种机制中，客户的经纪商可以自己作为客户委托单的交易对手方。经纪商这么做需要满足一定的条件，比如经纪商提供的价格不得劣于客户在公开市场上所能够获得的价格（在本章稍后的部分，我们将延续此话题讨论全美国最优买卖报价，即 NBBO）。然而，在证券交易所，交易者公开报出买入价格和买入数量进行申买（bid）并且 / 或者报出卖出价格和卖出数量进行申卖（offer）。直到有一方的申买与另一方的申卖相匹配——一桩交易就达成了。不同交易所之间的匹配撮合规则会有一点点差异，但其理念基本上都是一样的。

当前，一家典型的交易所不会只交易一只证券，而是交易很多证券。例如，在纽约证券交易所，交易者可以交易大约 3 000 只不同的股票。在芝加哥期权交易所，有接近 30 万份期权合约可供选择。作为公开聚集并且进行多种商品交易的地方，交易所与古代的集市并没有太大区别。实际上，据说集市这个词起源于波斯语 baha-char，或者叫“价格之地”。这听起来就像是一个交易所。

⊖ 除了私下寻找交易对手并且谈判成交之外，场外市场与交易所市场还有几点重要区别。比如，场外市场产品的标准化程度更低，没有中央清算机构，保证交易对手履约（例如，交易之后履行承诺）的手段相对较弱等，而以上这些对于交易所市场来说基本上是理所当然的。这些区分并不是仅仅只有纸面上的意义。2008 年，很大程度上正是由于 OTC 市场与抵押贷款相关的衍生品（如信用违约互换等）的大范围违约，才导致了美国自大萧条以来最严重的一次经济危机。

⊖ 纳斯达克曾经是一个 OTC 市场，并且现在也偶尔有人称之为 OTC 市场。但是这个划分在今天看来是相当脱离实际的，它并不符合我们这里有关 OTC 的说法。在我们的定义中，纳斯达克是一个交易所。

股票、期权和期货交易

表 2-1 列出了当前美国的四大股票交易所（按字母顺序排列），美国的绝大多数股票交易都发生在这 4 家交易所。表中还列出了这些交易所撮合引擎（matching engine）所在的地理位置。撮合引擎是实体存在的计算机服务器（或者更可能是计算机服务器组），每当潜在买家和潜在卖家在价格和交易量上达成一致，撮合引擎就实际进行配对撮合。

表 2-1 美国的股票交易所

交易所	撮合引擎
BATS	新泽西州威霍肯
Direct Edge	新泽西州泽西城（2010：新泽西州斯考克斯）
纳斯达克	新泽西州卡特里特
纽约证券交易所	新泽西州威霍肯（2010：新泽西州莫瓦）

撮合引擎已经取代了公开喊价交易池成为交易实际发生的唯一场所，在接下来讲到委托单簿（order book）时我们会再回头说到这个话题。由此可以推知，离撮合引擎越近，就能更快地向交易所发送和从交易所得到信息。因此，这些好斗的高频交易公司会尽量把他们自己的计算机放置的离撮合引擎更近一点，最好是能放在同一个数据中心之内，这种行为在实践上称为共置到交易所。

共置与传统的交易大厅中每天上演的一种常见策略非常相似。例如，直至今天，芝加哥期权交易所的 SPX[⊖]交易池中，交易员仍然在抢占距离活跃经纪商最近的位置。当经纪商从耳机里面得到一笔大额委托单报入时，那些挤得最近，甚至有时都踩到经纪商脚趾头的交易员比 5 个身位之外的交易员更有可能获得成交。这些交易员推挤得非常厉害，甚至连举起胳膊或者出去方便一下都几乎成了不可

⊖ SPX 是一种流动性非常好的指数期权，其价值与标普 500 股票指数当前的点位挂钩。

能之事。在交易池里面抢到一个好位置的益处已经众人皆知，毫不奇怪高频交易商在决定把它们的服务器放在哪个地方的时候也会有类似的想法。你有没有注意到 4 个主要的股票交易所都把撮合引擎放到新泽西？看来这个花园之州在荣誉称号之外[⊖]，还可以加冕世界共置之都的桂冠了。

表 2-2 列出了美国现有的和计划中的期权交易所。值得注意的是，CME 和 ICE 交易的是期货期权，而不是股票期权或者 ETF 期权，并且期权所基于的指数是在其他地方交易的。一份指数期货期权合约赋予你以一定价格买入或者卖出某期货合约的权利（而非义务）。期权到期时，并不是以现金或者股票进行交割，而是以股票指数期货合约进行交割。

表 2-2 美国的期权交易所

交易所	撮合引擎
BATS 期权	新泽西州威霍肯
波士顿期权交易所（BOX）	新泽西州纽瓦克（2010：新泽西州斯考克斯）
芝加哥期权交易所（CBOE）	伊利诺伊州芝加哥
CBOE C2（2010）	新泽西州斯考克斯
芝加哥商品交易所（CME）	伊利诺伊州芝加哥
洲际交易所（ICE）	伊利诺伊州芝加哥
国际证券交易所（ISE）	泽西城（2010：新泽西州斯考克斯）
纳斯达克期权市场（NOM）	新泽西州卡特里特
纳斯达克费城期权交易所（NASDAQ PHLX）	新泽西州卡特里特
纽交所全美交易所（NYSE AMEX）	新泽西州威霍肯（2010：新泽西州莫瓦）
纽交所群岛交易所（NYSE Arca）	新泽西州威霍肯（2010：新泽西州莫瓦）

最后作为本节的结束，表 2-3 列出了目前美国主要的股指期货交

⊖ 作为布鲁斯·斯普林斯汀（Bruce Springsteen）的故乡，该州常常受到喜剧演员的嘲弄，并且该州还是仅有的两个你仍然不能自己给汽车加燃气的州之一，另外一个 是俄勒冈州。据我所知，俄勒冈州还没有任何的共置设施。

易所。请注意在这些交易所也可以交易期货期权[Ⓐ]。

表 2-3 美国的股指期货交易所

交易所	撮合引擎
芝加哥商品交易所（CME）	伊利诺伊州芝加哥
洲际交易所（ICE）	伊利诺伊州芝加哥

另类交易系统

曾经，你可以在任何地方交易一只受欢迎的美国股票，只要这个地方是纽约证券交易所。[Ⓑ] 或者最起码看起来如此。在 20 世纪晚期，纽约证交所的行情显示牌就开始失去其垄断地位，到了 21 世纪前几年，这个趋势甚至有加速之势。直到最近的 2005 年，纽交所还可以自夸说在其上市的股票 80% 的交易仍然是在纽交所进行的，但到了 2009 年，这个份额缩减至仅大约 25%。[Ⓒ] 这些交易量去了哪里呢？大部分都是转移到证监会称之为另类交易系统（alternative trading systems, ATS）之类的平台里面去了。另类交易系统是一个包罗万象的概念，它包括所有将潜在的股票买家和卖家进行撮合，但又不是交易所的地方（期权和期货的大部分交易仍然是在传统交易所进行的）。

ATS 这个词过去指的是交易场所用电子化的方式将买方和卖方配对，而不是像传统交易所那样在交易大厅进行人工交易。但鉴于现在所有的交易所都是电子化配对了，这个区别现在来讲最少也是跟不上时代了。ECN，或者叫电子通信网络，就是一类非常盛行

Ⓐ 不管你是出于什么原因仔细阅读了这几页，你都可能会想了解如下事实：出于某些无厘头的原因，美国的期货交易所并不像股票和期权市场一样，由证监会监管，而是由美国商品期货交易委员会（Commodity Futures Trading Commission）监管。

Ⓑ 当然，这是我厚脸皮从亨利·福特那里借用来的概念（福特曾经说过：客户可以把他的车漆成任何颜色，只要这个颜色是黑色。——译者注）。

Ⓒ 美国证券交易委员会，“股票市场结构概念公告”（Concept Release on Equity Market Structure），2010-1-14[版本编号 34-61358；文档编号 S7-02-10] “2005 年 1 月，NYSE 执行了其上市股票交易量的大约 79.1%，2009 年 10 月，这个数字为 25.1%。”

18 打开高频交易的黑箱

All About High-Frequency Trading

的 ATS，其功能和交易所很像，不过它并不直接报价，而是通过纳斯达克进行报价。而暗池（dark pool）则是不发布，或者不显示任何报价[⊖]。一个交易场所可能在开始创建的时候是一个 ATS（例如 BATS），而随后就转变成了交易所。

很多概念很容易混淆在一起，比如交易是在哪“打印”的（交易公布），以及在哪“清算”的（交易实现）等很多诸如此类的东西，所幸的是，目前这些概念并不影响我们的进一步讨论。除非特别指出，本书所涉及的所有交易均应认为是在交易所发生的。并且，现在大多数的交易确实仍然是在交易所进行的[⊖]。在我们看来，这些交易场所背后的核心其实是相似的，这也就是其中真正重要的东西——委托单簿。

委托单簿

在古波斯的集市上，人们按照买卖意愿聚集在不同的地方：卖毯子的在这儿，卖羊的在那儿，卖驼绒毯的在橄榄树底下。在每个货摊之前，潜在的买家和卖家进行讨价还价，直到他们达成协议。之后，买卖双方可能会离开谈判地点去真正执行这笔交易，用商品去换取钱币或洋葱或其他什么东西——如果愿意，你可以把这个过程叫做清算。在现代的证券交易所里，买卖双方主要是电子化撮合撮合的，双方的讨价还价过程是在一个叫做委托单簿的地方进行的。所有的交易所和 ATS 都使用委托单簿，并且所有委托单簿的工作方

⊖ 后文我们会谈到储备委托单（reserve order），这种委托单可以撮合成交但是交易所不予显示。如果交易所所有的委托单和报价都用这种方式处理的话，那么暗池和交易所的区别也就不大了。

⊖ Direct Edge 最近由 ECN 转为交易所，证监会于 2010 年 3 月授予其交易所的地位。

式也都大同小异。委托单簿就好比核反应堆里面的控制棒，是一切反应发生的地方。

交易所的每一只证券都必定会有一个对应的委托单簿。对于某只给定的证券，比如 Acme Explosives 公司（代码 AEX），委托单簿看起来会像是下面这个样子：

500		1.00 × 1.10		600
-----	--	-------------	--	-----

位于中间的价格数据分别是**申买价**和**申卖价**[⊖]（申卖价既可以说是 ask price，也可以说是 offer price）。申买价是某方，或者是某几方想要买入的价格。申卖价是有人想要卖出的价格。如果你想买，你就得参考申卖价。想卖的话，就得参考申买价。位于外侧的数字分别是在这些价格上能够交易的数量。不是想买一些股票吗？有人愿意一美元外加 10 分钱的价格卖出 600 股。想卖？有人愿意 1 美元的价格买 500 股。顺便提醒注意，报价人的身份信息是没有公布的。只有交易所才知道是谁在进行买入申报和卖出申报。

以上这 4 个数字，只要是想在交易所交易的投资者，随时都可以看得到。这 4 个数字合在一起统称 BBO，即**最优买卖报价**（best bid and offer）。初看起来这些东西都挺没意思的。但至少有两件事情，让这些数字变得非常有趣。一是 BBO 是如何随时间发生变化的，比如在一个重要消息公布之后，或者刚刚发生了一笔大规模交易，供给需求规律大发神威（不只是一点点）。另一件有意思的事情是没有显示出来的交易意向。比如，可能有一个潜在的买家愿意支付的价格高于 1 美元，或者一个卖家愿意卖出的价格低于 1.01 美元，但是出于种种原因，他们不愿意把他们的买卖意向公布出来。有关 BBO

⊖ bid price / ask price（即 offer price），中文可译为申买价 / 申卖价，也可以依上下文译为买入价 / 卖出价、买价 / 卖价、买入报价 / 卖出报价等。——译者注。

的动力学，以及隐藏流动性（hidden liquidity）产生的效应，会在本书中的剩余部分占据很大的篇幅。这也是有关高频交易种种众说纷纭的议题所在。

除了 BBO 或者叫**单簿顶部**（top of book）之外，委托单簿还有很多其他内容。委托单簿中还保留有不那么好的买价（not-quite-best bids）和不那么好的卖价（not-quite-best-offers），如下所示：

...	600		0.98	550		0.99	500		1.00	×	1.10		600	1.11		800	1.12		950	...
-----	-----	--	------	-----	--	------	-----	--	------	---	------	--	-----	------	--	-----	------	--	-----	-----

这里我们看到的就是所谓的**深度委托单簿**（depth of book）的一部分。在 BBO 的左边，是低于 1.00 美元的一系列依次降低的申买价以及对应的申买量，右边则是高于 1.10 美元的一系列依次升高的申卖价以及对应的申卖量。这些不在场上的报价坐在委托单簿的两翼，就像在等待上场机会一样。就拿上面例子中的市场来说，比如，有一个卖家到来并卖出 500 份证券给报价为 1.00 美元的申买者，那么当前的报价将被从市场中移走，并且最优申买价会变成 99 美分的价格申买 550 股（顺便提一下，以申买价卖出称为吃掉申买，以申卖价买入称为打掉申卖^①）。如果该卖家不止卖出 500 股，比如说卖出 600 股，那么这 600 股将吃掉所有在 1.00 美元上的申买（有一个形象的说法叫**打穿**），并且还会有 100 股和 0.99 美元上的申买者成交^{②③}。

① 以申卖价买入又称为 taking the offer。（中国证券和期货市场暂时没有做市商制度，汉语中似乎也缺乏与 bid、ask、offer 等意义完全对应的词汇。以申买价卖出和以申卖价买入的行为在国内有人称之为“对价成交”。——译者注。）

② 这种现象有时称为**走过委托单簿**（walking the book）。有些交易所对此采用不一样的处理方法，例如将剩余委托单转变成限价委托单等。但在本书中，我们假设交易所采用的是这种机制。

③ 为了方便说明，这里还假设 Acme 的股票只在这家交易所交易。在美国市场上，如果 Acme 的股票是在多家交易所交易的，该交易所需要按照全美市场系统修正案（Regulation National Market System）的要求，去其他交易所寻找 1.00 美元的申买价，如果其他交易所没有更优报价，才能以 0.99 美元成交剩下的 100 股——除非该交易所将委托单进行闪现（这里还来不及谈到闪现的话题，但后文会进行讨论）。

NBBO

大部分证券都不止在一个交易所交易，因此，在任一时刻，都同时存在好几个活跃的委托单簿，每个交易所的每一只股票都各有一个。如前所述，每个交易所都有自己的最优买卖报价（BBO），这些报价有时是相同的，有时是不同的。而全国最优买卖报价，或者叫 NBBO，指的就是所有交易所^①中最高申买价和最低申卖价。下面的例子给出了 3 个 BBO 和由此导出的 NBBO：

交易所 A	"AEX" BBO 500 1.00 × 1.09 400
交易所 B	"AEX" BBO 200 1.02 × 1.11 600
交易所 C	"AEX" BBO <u>700</u> <u>1.01</u> × 1.09 300
	NBBO 200 1.02 × 1.09 700

2005 年全美市场系统修正案生效之后，NBBO 成为美国市场中一项非常重要的规则。该规则规定（还有其他一些规则，但这条规则最为重要），每个公众投资者都有权以 NBBO 进行交易，而不管该投资者将委托单递交给哪个交易所。

市场价差

申买价和对应的申卖价之间的差称为价差（spread）^②，又称为市场宽度（market width），价差正中间的理论价格称为市场中间价^③，或简称为中间价。价差（还有申买量、申卖量等，也称为深度）常常用来指示市场的流动性。流动性一直没有精确的定义，但当人们看到它时基本上都还是认得出来的。在流动性比较好的市场上，货

① 这里我们所称的交易所也包含 ATS。很重要的一点是，由于暗池（定义上）不发布报价，因此一般计算 NBBO 的时候也就不对其进行考虑。
② 注意不要和价差交易（spread trade）搞混了，本书不谈价差交易。
③ 为了精确计算中间价，你只需把申买价和申卖价相加再除以二。当然，你可能会得到一个分数的结果。

物比流动性差的市场上“流动”得更容易，或者说交易得更多。这是因为不论是对于买家还是卖家，流动性好的市场上价格都要更好一点（在这些价位上能够进行交易的商品也很多）。这里的关键之处在于，卖家喜欢更高的申买价，而买家喜欢更低的申卖价，因此较紧的价差对买卖双方都有好处。

考虑两个假设的市场都交易某只证券：

5 000		1.00 × 1.13		5 000
5 000		1.06 × 1.07		5 000

这两个市场的中间价都是 1.065 美元（ $[1.00+1.03]/2$ ）。然而，在第一个市场中买入 5 000 份证券需要支付 5 650 美元，而在第二个市场中只需支付 5 350 美元。同样，在第一个市场中卖出 5 000 份证券只能够收到 5 000 美元，但第二个市场中能收到 5 300 美元。很明显更紧的市场对买家和卖家都是更好的。在历史上，美国股票市场采用十进制之前，最小的市场宽度是 1 美元的 1/8，即 12.5 美分，那时的市场很像这里的第一个市场，不论是买家还是卖家都深受其苦。现在的股票（还有越来越多的期权）都按分计价[⊖]，市场上大部分比较受欢迎的证券都像是第二种情况。据说高频交易者在增进市场流动性方面起到了至关重要的作用，我们回头会再来审视这个说法。

价格动力学

在后文中，我们讨论交易委托单的市场影响，或者说一个委托单，尤其是大规模委托单的出现是如何驱动市场价格“远离委托单”（away from the order）的。如果有大规模委托单（有时称为

⊖ 有些市场允许小于 1 美分的最小变动价位。

block[⊖]，或者是 size order）卖出，价格会趋于下降；如果有委托单买入，则价格会趋于上升。这种现象对交易而言至关重要，有经验的交易员凭直觉也会知道这一点，甚至可以说，没有这种现象，就没有高频交易。所有这些都是供需关系决定的。在非金融市场，供求关系规律十分明显和直观。某种商品的需求增加了，由于买家之间的相互竞争，其价格会上升。因此，需求与价格有一种正相关的关系。相反，供给与价格则是负相关的。当供给增加时，由于卖家之间的相互竞争，价格会下降。我们对这种现象太习以为常了，以至于根本不会去想这件事情。拿汽油来说，夏天汽车涌向公路的时候，油价可能会涨，而油企生产过度时，油价又会下降。

买方的压力促使价格升高，卖方的压力促使价格下降。我们在证券交易所看到的正是这种现象。申买量的增加会推高价格——包括申买价和申卖价。申卖量的增加有相同效果，只不过方向相反。可参见表 2-4 的例子。

表 2-4 市场动力学

时间 1	500 1.00 × 1.10 500	市场处于平衡状态，中间价为 1.05
时间 2	2 500 1.00 × 1.10 500	买入申报增加了 2 000
时间 3	2 500 1.02 × 1.12 500	市场进行再平衡，中间价达到 1.07
时间 4	2 500 1.02 × 1.12 2 500	卖出申报增加了 2 000
时间 5	2 500 1.00 × 1.10 2 500	市场进行再平衡，中间价达到 1.05

后文中我们将更加精确地讨论申买和申卖数量增加对价格的影响，但在这里，我们只需要了解买卖动力关系同样会影响 BBO，这与影响汽油、苹果、MP3 播放器等的价格并无二致。需求增加推动价格上升，而供给增加推动价格下降。

⊖ 有关什么是大额交易（block trade）没有一个很好的、广泛接受的定义。曾经单笔 10 000 股以上的委托单称为大额委托（block），但实际上委托规模有多大与股价是密切相关的。如果一只股票价格只有几分钱，交易 1 万股算不了什么，而交易 10 000 股谷歌的股票则是另一回事了（在写本书的时候，谷歌股价在每股 500 美元之上）。

隐藏流动性

在前面的例子中，委托单规模的改变每个人都能看到，价格也如人们所预期的那样会发生改变。当需求增加时，卖方会提升卖出报价，因为他们清楚地知道，市场能够承受这一点。这种行为反过来又会促使买方愿意支付更高的价格。而如果卖出委托单规模增加，情形则会完全相反。

想象一下，如果前文的那个想在 1.00 美元价位买入 2 000 股的买家没有提交买入申报，而是在推特上同时告诉其他所有人他即将买入。那么你想价格会不会涨呢？当然会涨。此时卖家得到的市场信号和之前是一样的，即市场能够承受更高的价格，只不过这个信号来源是推特，而不是委托单簿。如果有人想以申卖价卖出 2 000 股，他没有在委托单簿中进行卖出申报，而是通过其他什么手段告诉其他人这个想法，其所引起的市场效果也是一样的。

这里要表明的观点是，如果你获悉了流动性的改变，尤其是那些有显著影响的改变，这种信息是市场供需力量变化的一个信号，它要么会像第一种情况那样以可见的方式影响价格，要么像第二种情况那样以隐藏的方式影响价格[⊖]。

交易

现在我们已经了解了各种类型的证券以及其交易场所，还了解了非常重要的委托单簿，现在让我来谈一谈交易行为本身。交易行为从递交一张委托单开始。对我们而言，一张委托单包含如下四项基本要素：

⊖ 在下一章中，我们会讨论储备委托单和冰山单，这是另外两种隐藏流动性的方式。

- 证券代码
- 买 / 卖
- 数量（规模）
- 价格

有时候，如何口述一份交易指令有非常成型的惯例。如果你想以 1.00 美元每股的价格买入 500 股，你应该说“一块买五百”。如果你想卖，你应该说“五百卖一块”。[⊖] 随着交易自动化程度的提高，这些东西也就渐渐变得没有什么用途。不过这倒是一个有意思的语义效率的范例，最起码对那些对语义效率范例之类事情感兴趣的人很有意思，不过这句话本身实在算不上语义效率的范例。

在递交一张委托单时，除了可以指定基本要素（证券、买卖、价格、数量），你还可以指定委托单的处理方式，其中最重要的是指定该委托单是一张市价委托单（你将得到市场目前最优的卖价或者最优买价）还是限价委托单（你指定想成交的价格）。对于限价委托单，你还可以指定交易所在多长的时间段内试图执行该委托。最为常见的限定有取消前有效（good-till-cancel，顾名思义）、当日有效（good-till-day，如果交易收盘时还未成交则取消该指令）、不成即撤（immediate-or-cancel，立即执行能够成交的部分，没有成交的全部撤销掉）和全部全不（all-or-none），或者全成全撤（fill-or-kill，如果不能立即全部成交，则撤销委托）。有些交易所对这些限定有微小的修改，或者采用其他名称来表示这些限定，但是委托单最本质的东西——表明买入意向或者卖出意向——在哪里都是一样的。

对于股票而言，我们最好区分一下两种基本的卖出指令，分别

⊖ 一块买五百，这句话的英文原文是“one dollar bid for five hundred (price for size)”，五百卖一块，英文原文是“five hundred offered at one dollar (size at price)”。——译者注

是多头卖出（long sells，我拥有某只股票并且我想卖出它）和空头卖出[⊖]（short sells，我借入股票并且卖出，将来我会将股票买回以归还给借出者）。这两个概念的区别常常让人感到困扰，不过如果理解了交易者这么做的原因，其区别也就显而易见了。多头卖出某只证券，是因为他出于某种原因，不想继续持有该证券。空头卖出某只证券，是因为他想从价格下跌中获利。举个例子：我找你借入股票，并且以 10 美元的价格在市场上卖出，接下来该股票价格下跌至 9 美元，这时我就能以 9 美元的价格将股票买回还给你，自己则从中获取 1 美元的利润；当然，如果市场价格上涨，这笔交易我就赔钱了。接下来我们要谈到的期权的做市商就是非常活跃的股票空头卖出者。他们卖空股票并不是为了投机，而是为了锁定卖出或者买入期权合约所获得的收益，也就是给他们的期权头寸做对冲。

2008 年市场崩盘的时候，卖空行为引起了巨大的争议。批评者认为卖空加速了市场下跌。他们希望重新引入所谓的“提高报价规则”[⊕]，只在价格有上升时，才允许空头卖出，从而避免卖空者在股价下跌时落井下石。而卖空行为的支持者则认为卖空是投资者的一项基本工具，只有这样才能保证市场价格真实反映股票价值。如果投资者认为相对于公司的基本面——比如潜在的盈利能力、管理的优良程度、市场地位等——股价低估了的话，他们可以买入股票。如我们之前所说的，买入力量增强将导致股价上涨，并且直到投资者认可的程度，从而使得这些投资者获得一笔收益。为了获得这些收益，投资者也有动力努力寻找此类投资机会。相应的（如果你是

⊖ short sell 后文也译为“卖空”。——译者注

⊕ “提高报价规则”，即 uptick rule，指的是卖空者只能在两种情况下卖出，一是卖出价比上一笔成交价高，二是卖出价和上一笔成交价相等，但是上一笔成交价要比上上笔的成交价高。——译者注

卖空行为的支持者，这一点至关重要），如果投资者认为股价过高，也应该能通过另一个方向纠正股价而获利。

还有一个概念在高频交易中经常出现，即被动交易（passive trading）和主动交易（active trading）。被动交易是指你提交一张委托单，而不确定是否会有人成为你的对手方，而主动交易则是提交一张委托单直接和市场上显示的申买或者申卖成交。比如对于 1.00×1.10 的市场，如果你进入这个市场，提交了一张 1 元的申买委托，或者 1.1 元的申卖委托，那么你就是被动交易。即使你提交 1.01 元，甚至 1.09 元的申买委托，你都还是被动交易，因为市场上显示没人有意向以低于 1.1 元的价格卖出。而这个市场中的主动交易的例子则是直接吃掉 1 元的买价或者打掉 1.1 元的卖价。在这两种情况下，你都知道有人愿意以这些价格交易。你说，好吧，我来做交易的对手方。被动交易也叫做提供流动性（making liquidity），主动交易也叫做取出流动性（taking liquidity）。

当你进行主动交易时（这里假设你能成交，因为可能会有人在你之前把你想成交的委托单全部吃掉），你从市场中移除了流动性，或者叫把流动性取走了，所以主动交易有时又称为取出交易（take-out trading）。高频交易公司会使用专门的软件程序进行这类交易，因此这种程序又名取出引擎（take-out engine）或者电子眼（electronic eye）。进行被动交易的程序又称报价引擎（quoting engine）、自动报价机（autoquoter），或者流式报价机（quote streamer）。

交易者

虽然交易所内做的事情只有一件，就是交易，但人们参与交易

的原因却是多种多样的。你参与交易的原因会深刻地影响到你的交易习惯，也即你提交申买和申卖委托时的具体行为——你报出的价格和数量，被动交易还是主动交易，你如何指定委托单的细节，什么时候撤销或者修改委托单，等等。这些交易习惯通常称为策略（strategy），接下来我们会再谈到这一点。不过在此之前，为了更好地讨论这个涵盖面很广的话题，我们先区分4种类型的交易者：投资者（investor）、做市商（market-maker）、套利者（arbitrageur）和预测者（predictor）。这些区分都仅仅是个原型，如果你愿意，你可以将这些概念进行大范围的泛化。实际上，当然不太可能有哪个交易者或者交易公司一定属于并且只属于其中的某一类。一家公司可能会顶着不止一个头衔——这往往取决于那些冗长的规则和监管制度。不过，这些原型仍然会有助于我们理解后文所要谈到的交易策略。

投资者

投资者进行交易通常是因为他们想增加或者降低在某只证券上的仓位。问题的关键在于，投资者增加或者降低某只证券的持仓，是出于投资证券的本质利益。他们买入一只股票，是因为他们认为其价值会上涨，卖出股票是因为他们认为其价值会下跌，或者，他们卖掉投资组合是认为股票价值不会再上涨或者已经涨到了理想价位。当投资者持有一只股票的时候，他们倾向于持有一段较长的时间——比如几周、数月或者几年。当然，你可以把这些标准定得更细致一点。对于那些进行非常短期（比如，几天、几分钟、几秒钟或者几毫秒）“投资”的个人或者公司，我们把他们归为预测者这一类，我们会在后文对此进行讨论。

投资者进行交易的时候，他们自然想获得最优的价格——当卖出

的时候，他们想得到最高的申买价，买入的时候想得到最低的申卖价。对于投资者而言，较紧的市场要优于较宽的市场。并且，投资者当然还希望能把交易费用（例如交易所费用和经纪费用）降到最低。

在投资者这个类别中，既有个人投资者，也有机构投资者。前者一般是指某直接交易的个人，他通过在线经纪服务之类的途径进行交易。后者则是代表一群人进行交易。这一群人可能是共同基金或者对冲基金的投资者、养老基金的未来受益人，或者是其他一些大型组织。这里的关键词是大。机构投资者的委托单规模常常比个人投资者大很多。另外还有一个概念与机构投资者紧密联系，那就是他们对投资者的信托责任，即他们有义务以对投资者最有利的方式进行交易。这种信托责任，加上他们交易的委托单的巨大规模，使他们对委托单和交易的市场影响非常敏感（前文我们讨论价格动力学和隐藏流动性时对此有过叙述）。当我们对投资者的交易策略进行探讨时，这种敏感性就特别明显。

一般而言，以投资者身份或者代表投资者利益进行交易的交易员称为买方交易员。这个称呼与字面意义不太相符，因为投资者并没有什么偏好一定要去买入，他们同样也会卖出。不管怎么说，这个称呼还是沿用了下来。卖方交易员这个词常常指的是做市商（马上就说到他们了）之类的交易员，不过他们也并不是纯粹卖出，而是有买有卖。

做市商

在一个理想的市场中，投资者总是可以和其他投资者进行交易。有人想在投资组合中加入一定数量的股票，他总是可以找到那些想在投资组合中卖出相同数量股票的人，反之也是如此。实际上，在

有些市场，比如，对于那些交易最活跃的股票，有时确实能遇到这种情况。但在股票大超市的大部分地方，自然产生的流动性都不足以将一个投资者和其他投资者有效地配对起来[⊖]。这个时候，做市商就要登场了。

这些交易者，有时也称为专家（specialist），他们总是准备着按事先指定（或报出）的申买价和申卖价进行证券买卖。他们的存在保证了投资者在想要买卖的时候总是能够找到交易对手。如果观察一下委托单簿，你可以看到其中有些数量的委托单来自于投资者，有些来自于做市商。比如，在下面所示的 BBO 中，1.10 美元这个价位上可能有 500 份申卖委托来自于做市商，剩下的 100 份来自于投资者。对潜在的买家而言，他并不在乎卖方是做市商还是其他投资者。对他而言真正重要的事情是价格和数量。不难想象，有时候所有的申买委托和申卖委托都是做市商提供的。想象一下，如果 1 美元这个价位上的 500 份申买委托全部来自于做市商，而 99 美分这个价位上的 550 份申买仅有一部分来自于做市商，那么在这种情况下，做市商的出现就改进了市场上的买入报价。

...	600		0.98	550		0.99	500		1.00	×	1.10		600	1.11		800	1.12		950	...
-----	-----	--	------	-----	--	------	-----	--	------	---	------	--	-----	------	--	-----	------	--	-----	-----

一个纯粹的做市商没有任何兴趣持有股票。他们就是俗话说的行商，不是坐商。做市商的目的说来简单：他们就是想在一个价位上买进来，再在高一点的价位卖出去，做一轮转手，赚取中间的价差。因此，做市商的工作就是赚取价差——价差越大越好。当然，这与投资者的想法是恰恰相反的，对投资者而言，市场越紧

⊖ 对于期权和不是那么活跃的股票，情况肯定是这里所说的样子。但对于期货市场，尤其是我们接下来要谈到的期货市场，往往天然就有充足的流动性，这个时候做市行为实际上就没什么必要了。

越好。

如果做市商总是可以将自己的买卖交易配对抵消，低买高卖，那么就没有一点风险。但是这显然不符合事实。考虑一个紧致（并非完全不现实）的市场：

500		1.04 × 1.06		100
-----	--	-------------	--	-----

比如肯是一个做市商，他同时在买卖双方报价。有人和他的申卖委托成交了 100 股，他也因此以 1.06 美元的价格卖空了该股票。此时原本的最优申卖已经全数耗尽了，假设现在 BBO 变成如下这个样子：

500		1.04 × 1.07		500
-----	--	-------------	--	-----

他期待有人吃掉他在 1.04 美元的买入报价，这样他就可以平掉他的空头头寸，并且赚到 2 美元的价差。可是，当他正在等的时候，市场又变成了下面的情况：

500		1.07 × 1.08		500
-----	--	-------------	--	-----

咣当！现在他发现自己卖出了 100 股，但很难再以更便宜的价格买回来了。现在他在这个市场中能做的最好的事情是以 1.07 美元的价格买入股票，认赔 1 美元（100 股，每股赔 1 美分）。为了避免遭遇更大的风险，他不得不狠了狠心提高了卖出报价，这也显示做市这个事情本身是有风险的。

500		1.04 × 1.06		100
-----	--	-------------	--	-----

再举另外一个例子，这次的结局就要好多了。假设开始的时候市场情形是一样的：

500		1.04 × 1.06		100
-----	--	-------------	--	-----

安也是一个做市商，她同时在买卖双方报价。有人和她的申卖委托成交了 100 股，她也因此以 1.06 美元的价格卖空了该股票。此时 BBO 的变动和肯的情况是一样的：

500		1.04 × 1.07		500
-----	--	-------------	--	-----

安知道，如果有人打掉她的买入报价，她就可以赚到 2 美元。然而，安并不想冒这个险，因此她将买入报价提高了 1 美分。现在市场看起来是下面这个样子：

100		1.05 × 1.07		500
-----	--	-------------	--	-----

这时有卖家看到这个报价并且抢先出手，吃掉了她的买入报价并且以 1.05 美元的价格卖给她 100 股。她则把这些股票还给借给她的人，并且平掉空头头寸。因为她是以 1.06 美元卖出，1.05 美元买入的，因此赚到了 1 美元的利润。

这个例子说明了做市商所承担的风险，以及因补偿风险而赚取的买卖价差。如果我们把这个例子稍稍做一下扩展，就可以说明另一个动力学问题。比如还有一个做市商名字叫詹森，1.07 美元这个价位的 500 张申卖委托中，有一部分是他的。这时他看到安提高了申买价，决定跟随安的报价，再报出 500 股的申买单。此时的市场看起来是下面这个样子（注意安的 100 股买入委托已经成交了）：

500		1.05 × 1.07		500
-----	--	-------------	--	-----

现在一个投资者以 1.05 美元的价位卖出的委托单进来了，打掉了詹森申买委托中的 100 股。就像安那样，詹森也想尽快平掉仓位，因此他降低了当前市场的最优申卖价，以 1.06 美元的价位报出了 100 股。这时有人跟随他又报出了 400 股，就像他在安试图反手交易的时候也跟随了安一样。现在市场变成了以下这个样子：

$$400 | 1.05 \times 1.06 | 500$$

此时有买家觉得 1.06 美元的卖出价很有吸引力，他以这个价位买入 100 股。买家的委托单和詹森的申卖成交了（因为他排在前面），因此詹森以 1.06 美元的价位卖出了 1.05 美元买入的股票，并且赚到了 1 美元。注意到价差了吗？只有 1 美分了。而刚开始的时候，价差有 2 美分。因此，做市商之间的竞争迫使每个做市商都不得不接受一个更低的利润率，迫使他们不断地改进报价，最终的结果就是产生了一个对投资者而言更好的市场。

套利者

在物理世界之中，大质量的物体，比如行星，会将较小质量的物体拉向其质量中心，这就是强大且恒久的重力作用。重力一直存在，总是将空中的物体拉回到地面，不管是沙滩排球、棒球，还是黄油土司。在金融市场中，我们也发现了一个同样强大的力量，即一价定律。一价定律是套利定价理论的结果，这个理论使得一只证券的价格不会长远地偏离其真实价格。虽然真实价格会随着时间改变，但是在任一给定时刻，真实价格只能有一个。

这个定律的一个结果是：如果一只证券在多个市场交易，并且每个市场的交易成本都是相同的，那么在任一时刻，该证券在这些市场的价格都应该是相同的。如果价格不一致，套利交易者就会从低价的市场买入，并且在高价的市场卖出，从中获得无风险利润，直到他的交易行为使得市场价格回归一致。想象一只代码为 PDQ 的股票在两个交易所交易，这两个交易所的市场情况如下所示：

$$\begin{aligned} \text{交易所 A: PDQ } 500 &| 1.00 \times \mathbf{1.02} | 500 \\ \text{交易所 B: PDQ } 500 &| \mathbf{1.03} \times 1.04 | 500 \end{aligned}$$

这里我们看到一个所谓的交叉市场（crossed market），即其中一个交易所的申买价大于另一个交易所的申卖价。不考虑交易成本的话，一个套利者可以在 A 交易所以 1.02 美元买入，同时在 B 交易所以 1.03 美元卖出，他就可以因为发现了这个机会而进行交易瞬时取得每股 1 美分的收益。可能真实交易时，由于交易成本的关系，实际的利润会小于 1 美分，但不论如何，整个过程都还是类似的。当然，只要此机会存在，这个套利者会一直交易下去，一分钱一分钱的搬走利润，直到市场不再处于交叉状态。此时的市场可能看起来是下面的样子：

交易所 A: PDQ 500 | 1.00 × 1.03 | 500
交易所 B: PDQ 500 | 1.02 × 1.04 | 500

你可能已经猜到速度是进行这种套利的关键，因为谁不想进行这些交易呢？速度的重要性来源于两个方面。其一，有非常多的套利者在寻找这类错误定价，而能够最快速递交委托单的那个才能胜出；其二，如果两张委托单不是非常精确的同时发出和同时处理的，那么完全有可能在套利的两条“腿”还没有成交完毕的时候，某个市场价格已经发生了变动，错误定价不复存在，此时套利者只好持有单腿的仓位并且不得不找机会平仓出来。

股票大超市里挤满了套利者，他们寻找错误定价进行修正并获取利润——这些动作一定要尽可能的快。他们并不只是寻找与上面例子中类似的单只证券的错误定价，还在寻找价格有关联的完全不同类型的证券之间的错误定价。比如：ETF，或者指数跟踪股票池，其价格必须与它们所跟踪的指数一致。以组成标普 500 指数的 500 只股票为例，这个股票组合常称为标普 500 股票篮子。如果不考虑交易费用，Spider ETF 的价格应当永远是标普 500 的 1/10。如果代表一

篮子股票的指数的交易价格是 1 400 美元的话，那么 ETF 的价格应该是 140 美元。如果 ETF 的交易价格高于 140 美元，我们就可以卖出 10 份 ETF，并且买入股票篮子。如果交易价格低于 140 美元，我们就可以卖出股票篮子并且买入 10 份 ETF。在实际中，我们无法忽略交易费用和其他一些实际问题，因此 ETF 对指数篮子的套利会比这里看起来要复杂一些。不过因为有利可图，还是有很多套利者乐此不疲。

ETF 之所以要和股票篮子保持价格一致，是因为不论是 ETF，还是股票篮子，它们将来所产生的现金流都是一模一样的。这才是套利定价理论的关键。任何两个证券或者证券组合，只要其未来现金流一致，它们就具有相同的价格。这就让套利者在寻找套利机会时可以有一个非常有创造性的视角。例如，人们可以创建一个由两份股指期货合约组成的简单组合，其净盈亏与股指期货完全一致^①。如果这份合成的期货与真实的期货价格不一致，套利者就可以并且也愿意消除这个差异。

股票超市中不同证券的价格之间有一张微妙但是不可逾越的关系网，这张网巨大、复杂，并且稠密。随着时间的变化，这张网的形状和组成也在不停变化，但它总是把巨大的市场相连为一体。就像一只博德牧羊犬不让迷失的动物走得太远一样，这些套利者也非常积极，并且不知疲倦地将价格拉回到它们应在的位置。

预测者

前文所列举的套利的例子都涉及两个或多个证券之间的错误定

① 例如，有一种叫做“组合期货”(combo)的策略，该策略买入一份 SPX 看涨期权，卖出一份 SPX 看跌期权。这与买入一份交割价格和交割时间分别对应于“期货组合”执行价格和到期时间的 E-迷你标普股指期货合约是一致的。

价，当在同一个时间点寻找时，这些错误定价还比较明显。还有一类我们称之为预测者的交易者，你可以大致把他们想象成在不同时间点寻找价格差异。或许你已经预想到，这件事可就更难一些了。因为没有什么未来之事是绝对可以预测的，与简单套利相比，发现这些价格差异往往要不确定得多。

预测者，也叫做量化交易者，他们进行的不是我们可以称之为“静态”的套利，而是统计套利[⊖]。你可以找到很多不同的统计套利的定义，但它们都牵涉到分析历史数据、预测未来的价格方向、使用统计学和概率论的知识，并且依据这些预测做交易。如果你预测对的时候比错的时候多一些，甚至只多出一点点，你就能赚到相应的利润。

或许最常见的统计套利策略是所谓的配对交易。在这里，预测者找出两只价格倾向于走势相同的股票，比如可口可乐公司和百事可乐公司。你可能设想这两家公司的产品类似，因此其股价走势应该相近。如果你把这两只股票的历史价格曲线放在一起考察一下，会发现确实如此。预测者可能会分析这两条价格路径，并且非常精确地定量描绘出可口可乐和百事可乐股价之间的关系。有可能他们会发现，在一定的置信度之下，一只股票的价格变动会引起另一只股票在某个时间段之内也发生相应的变动，并且一只总是领先，另一只总是滞后。如果预测者看到领先的那只股票价格发生了跳升，她会立刻买入滞后的那只股票，期待其价格也会上升。如果价格真的上升了，他就可以卖掉股票并且获得收益。这里的关键词是如果，因为往往如此并不保证一定如此。

⊖ “静态”套利，原文是“static arbitrage”，统计套利则是“statistical arbitrage”，static和statistical这两个词形态相似，作者这里玩了一个小小的文字游戏。——译者注

人们研究配对交易已经很多年，有些人认为这个领域没有什么油水了。过去相关联的股票可能要花上几小时甚至几天才能跟上另一只股票，而现在配对交易者多如牛毛，股票价格之间的迟滞时间基本上短到没有。但如果考虑到股票超市中不同证券之间的关系网络，其中可能出现错误定价的次数却是相当多的。在下一章中，我们将就此进行更深一步的探讨，并且介绍一些预测者可能使用的利用这些定价错误的策略。

高频交易者

那么在对投资者粗枝大叶的分类之中，高频交易者应当归于哪一类呢？虽然这个问题合情合理，但是其答案却是难以捉摸，其中的原因也不是一两句话能够解释清楚的。首先，不论是哪个高频交易者，他都没有什么理性上的动机和你讨论这个事情。即使高频交易的那些赚钱传闻只有一点靠谱，你也可以想想，如果把时光机器调到1849年，你问问一个成功的淘金者，他会不会把他标记的地点告诉你，让你在周围挖上一阵子。实际上，高频交易者比1849年的那些淘金者还要保守秘密，这是因为他们的交易策略是非常容易移植的。在金融圣殿的这个角落之中，真正的秘密信息可以很容易地藏在一个知情员工的两个耳朵之间，整个圣殿的巨大空间都还没有用到呢。[⊖]

除了保护利润这个显而易见的动机之外，有可能还有些高频交易公司不想让其他人知道他们在掘金的时候都做了什么——即使所标记之处已经黄金散尽，他们也不想让他们的作为暴露在道德和法律

⊖ 比如说，2009年高频交易巨头城堡集团起诉了三个前雇员，理由是这三个前雇员违反了非竞争协议，创建了戴莎技术（Teza Technologies）高频交易公司。戴莎公司在此之前刚刚雇用了一个高盛前雇员，而这个雇员也被高盛起诉偷走了他们的高频交易机密，戴莎只好随后将此雇员炒掉。这一年对戴莎人来说还真是不错。

的强光之下。最后，但并非最次要，是我们可能没有一个很好的相关定义。高频交易者这个名词更多的是为行业外的人士发明出来的，和行业内的专业人士没什么太大关系。这也反映了人的一种天然需求，即用一个很简单的概念描述一个本质上十分复杂的世界，在这个世界里，可能一天中唯一简单的事就是泡壶咖啡了。

虽然定义上有些困难，但把高频交易者看成是拥有令人生畏技术的做市商和预测者的结合体，一般来说错不到哪儿去。这些卖方交易者与计算机兴起之前做市商的目标在本质上是一致的，即在买入价买入，在卖出价卖出，低买高卖，以赚取买卖价差。然而，由于十进制化的发展、计算机技术的进步以及越来越激烈的竞争等，如今的高频交易者相比传统的做市商不得不更加创新，更加激进，以及（有人会说）更多使用掠夺策略。在下一章中我们将介绍一些此类策略。相比于纯粹的做市商，高频交易者还更加具有选择性，在选择交易的证券时，他们偏爱那些执行速度能带来优势的，也就是说交易最为活跃的股票。[⊖]

高频交易者并不是唯一的想借助提升技术来获取优势的群体。在高频交易的语境中，你还会时不时地听到算法交易者（或称黑盒交易者）这个词，然而，在大部分情况下，这个词并非指的是高频交易者，而是买方中与高频交易者相对应的人。算法交易者更多的是一个机构投资者，其利用自动化的方法为客户服务。一般来说，算法交易者提升计算能力并不只是为了达成更快的执行速度（当然越来越多的人也在这么做），更多的是为了优化大型投资组合的配置、决定买卖股票的时机，以及最小化委托单的市场冲击成本等。

⊖ 你可能听说过高频交易者只交易最活跃的大约 300 只股票，但这个事情没有谁能说得准。有些高频交易公司交易股票数更少一点，有些则更多一点。

现在你已经知道两者的区别了，不过也不能对这里的定义抠得太严。我们定义的从事高频交易的公司和从事算法交易的公司当然有可能以其他我们谈到过的身份出现。比如，在最活跃的股票上进行高频交易的公司有可能还在不那么活跃的股票上进行做市交易，当然很有可能他们还有一套套利系统来抓取基本的套利机会。即使是算法交易者，他在设计算法时也可能会考虑预测者所使用的统计套利策略。

然而，一次交易中可以同时出现多少种身份有一个很强的限制，这就是所谓对客户交易和自营交易的隔离。这中间的理念是，一个经纪自营商在寻求以最优价格执行客户委托的同时，不能也出于自身利益为自己下达委托单，即人们所知的“为自己账户交易”。抢先客户进行交易具有很大的诱惑，最少客户得考虑一下他是否从经纪商那里得到了他可能得到的最好服务。只进行客户交易的公司，或者只进行自营交易的公司对这种隔离不必顾忌太多。但是这两种交易都有的公司（很多华尔街公司和对冲基金都是如此）就必须在客户部门和自营部门之间建立所谓的信息屏障上下工夫了。

第 3 章

All About High-Frequency Trading

交易策略

上一章中，我们根据交易者入市的原因将证券交易者分为 4 种类型，其中投资者持有证券的周期一般较长；做市商赚取买卖价差，承担双边报价的风险；套利者交易定价错误的证券，卖出被高估的同时买入被低估的，赚取之间的差价；预测者采用量化交易技术（或统计套利）分析数据，利用它来预测价格未来的变化，如果价格变化与预测一致，就能获利。高频交易者，如我们介绍的那样，本质上就是做市商和短期预测者的混合体。

在本章，我们将讨论一些交易者可能会出于种种目的而使用的短周期策略。使用短周期这个词，是要特别强调对于这类策略而言，时间是至关重要的因素。这里将不涉及任何长周期策略，比如投资什么类型的证券、持有多长时间、如何分散投资组合等。这些内容能在其他相关的书籍中找到。

那么这几页是否把所有赚钱交易者所使用的策略都讲到了？一条漏网之鱼都没有了吗？当然不是。要是能有这样的书就好了。正如你想的那样，市场上有比这多得多的交易策略。对于交易公司而言，交

42 打开高频交易的黑箱

All About High-Frequency Trading

易策略是他们的最高机密。就好比可口可乐的配方，或者足球比赛的战术手册一样——我想你懂的。这里讲的策略一般都是有一些年份了，以至于可以从一个公司流到另一个公司（交易员也会跳槽），或者人们会在公开的演讲中提到它们。其中有一些策略相当直观。你可以打赌，还有些策略就是用来玩爆它们的，那些薪酬不菲的量化分析师还在不停地想着新的策略。不过我们总是得有个开始，是吧？

通常，我们假设交易发生在充满大量主动和被动交易者的交易所里。我们还假设交易所采用价格-时间优先的撮合机制，委托单按照其到达交易所的先后顺序成交，并且报价较好的委托单（1.00 美元买入）比报价较差的委托单（0.90 美元买入）先成交。同时，我们还假设交易所采用流动性提供者-取出者定价模式，交易所会给流动性的提供者（挂出新的申买或者申卖委托）一定的返还，以吸引他们到交易所交易，而取出流动性的交易者（吃掉申买或者申卖委托）则需要支付一定的费用。举一个提供流动性的例子：市场当前报价是 1.00×1.10 ，弗里克先生挂出一份申买委托单，使市场变成了 1.01×1.10 。再举一个取出流动性的例子：弗兰克先生吃掉 1.01 价位上的买单并支付取出流动性的费用，比如一分钱。交易所把这一分钱分一部分给弗里克先生。

美国大部分股票交易所和一部分期权交易所采用的都是价格优先-时间优先和提供者-取出者定价模式（例如纽约群岛期权交易所和纳斯达克期权市场）。还有一种定价模型称为散户优先模式（例如芝加哥期权交易所和国际证券交易所），这种模式向专业交易者收费，而不向非专业交易者（或者散户）收费。这些期权交易所还会根据交易者所采用的身份，把散户委托单放在专业交易者的委托单之前成交。最后让这个图景更完整一点，再介绍一下美国最主要的

期货交易所（芝加哥商品交易所），在这个交易所，不管是提供者还是取出者，都需要交费，通常会员的交易费用会比非会员低一些。你可能会觉得收费返还这类政策都是小事，不值得一提，但实际上它们的影响非常之大。例如我们下面可以看到，在买卖价差小到基本上没有的情况下，返还可能是做市商唯一的利润来源。

投资者与做市商

我们首先考虑的两类交易者分别是机构投资者（也就是所谓的买方）和纯粹做市商（卖方），我们将考察他们各自的利益诉求和交易策略。实际上，这里所说的“投资者”很可能是代理真正投资者进行买卖的经纪自营商。为了简单起见，我们不打算讨论投资者和经纪商的关系，并且把后者从这个抽象的图景中剔除掉。我对我的经纪商朋友们没有丝毫不敬的意思，只是不讨论这些模糊的领域也不会影响到对高频交易实质的理解。我想他们会理解的。

之所以强调是“纯粹”做市商，我们指的是这家做市商仅仅赚取买卖价差或返还，而不采用套利策略。在大部分市场上，投资者和做市商都是唇齿相依的。投资者需要做市商提供流动性（例如，为他们的交易找到交易对手方），而做市商需要投资者成交他们的申买单和申卖单，以赚取买卖价差或者返还。^① 但是上一章已经讲过，到了买卖价差这个事情上，他们的偏好就截然不同了：做市商希望价差越大越好，而投资者希望越小越好。因此，投资者和做市商又是天然对立的。他们的关系虽然不像哈特菲尔德家族与麦考伊家族^② 那

① 在有些市场中，投资者所提供的自然流动性非常充裕，也就是说，投资者很容易找到与其方向相反的交易对手，这样他们就不是很依赖做市商了。

② 哈特菲尔德（Hatfield）家族与麦考伊（McCoy）家族是美国 19 世纪下半叶两个有世仇的家族。——译者注

样仇深似海，但也像小骗子麦吉与莫莉[⊖]那样龃龉不断。在每一个强健和充满竞争的市场上，投资者和做市商之间的强大利益冲突都使得他们各自为阵。

假定我们的投资者是一个为其委托人进行交易的机构投资者。虽然我们讨论的大部分东西对于提交小额委托单的个人投资者也是适用的，但是出于现实的原因，那些大额委托单（大宗交易）更加吸引我们的注意，因为这些委托单有可能使得市场价格朝不利于投资者的方向运动。这就引出了投资者最为关心的一点：如何完成交易（例如，找到流动性，找到交易对手），同时又将其委托单的市场冲击降至最低。她不希望她的委托单将市场推到对她不利的方向——或者，更准确地说，推到对其委托人不利的方向，她有责任在尽可能好的价位上为其委托人进行交易。这种责任也意味着投资者倾向于不支付买卖价差。我们假设她愿意以中间价成交，当然，她更乐意以最优买卖报价（以申买价买入，以申卖价卖出）成交，而不愿意支付买卖价差（以申卖价买入，以申买价卖出）。但是一般而言，中间价也是可以接受的。自然，投资者还希望能尽量降低交易成本，比如，降低交易所费用和经纪费用等。

相对于投资者而言，做市商策略的驱动因素就少一些，这当中最重要的一项就是风险管理。回想一下我们之前讲过的，做市商没有任何意愿长时间持有证券。对做市商来说最好的情况就是，他能以申买价从一方（比如一位投资者，不过他才不在乎那个人到底是谁）买过来，然后立刻以申卖价卖给下一个人，从而自己两手空空。反过来，他也想以申卖价卖给一位投资者，再立即反向从另一位那里

⊖ 小骗子麦吉与莫莉（Fibber McGee and Molly）是美国一部 1935 ~ 1959 年持续热播的广播喜剧。——译者注

以申买价买入，同样不留下一片云彩。要记住，做市商做的是流动业务，不是存储业务。但他的世界并非如此完美，在前面的章节中我们也看到，最坏的一种情况是做市商以一个价格进行交易，然后又不得不以更差的价格进行反向交易。例如，做市商还没来得及完成双向交易，市场价格就发生了对他不利的变动，迫使他不得不高买低卖。

由于做市商在一个交易来回之间总是会有时间差，并且这段时间之内价格的变动是难以预测的，因此，做市商一定会面临风险。你可能已经想到，由于双边报价承担了不可避免并且相当巨大的风险，做市商会想最大化其利润作为风险补偿。如果卖出了一份证券，他不只是期待以更低的价格买回，而是想以最低的价格买回。如果买入的话，他则会期盼以尽可能高的价格卖出。这并不只是为了实现利润最大化（当然他们想这样），而且还是为了对那些不可避免的坏交易进行补偿。这些交易会直接把钱从他们口袋里掏走，即使是最好的做市商，他们也时不时地会遇到这种情况。在这里失去的，他们希望能从其他交易中赚回来。

这两个驱动因素——避免不利交易和最大化风险补偿——使得做市商对前文中我们谈到的一个东西特别敏感，这就是对市场还不可见的大额委托单。投资者想尽可能长时间地隐藏其大额委托单，做市商却非常想发现这些委托单，以便避开它们；或者，在条件许可的情况下，抢在它们前面进行交易（在后面的策略部分我们将进行更多讨论）。

我们会在后文中专门讲到速度问题，但是这里有必要先说一说交易耗时对投资者和做市商各自的重要性。简而言之，这里所说的交易耗时指的是将委托单（对投资者而言）和报价（对做市商而言）

报入或者撤出市场所需要的时间。投资者和做市商都希望尽可能快的把他们的委托单和报价报入或者撤出市场，这背后的原因大致相同，他们都希望在有利价格消失前抓住它，在不利情况发生时尽快跑掉。不同之处在于交易涉及的数量。机构投资者可能一次仅提交一份委托单，或者是几份，也有可能到 100 份。但典型的做市商则会同时向市场提交好几百甚至几千份报价。一个期权做市商可能市场中有成千上万的报价，其中每份报价都是一份交易承诺。一个快速的意外市场波动对投资者来说是痛苦的，但对做市商来说却是灾难性的。投资者和做市商都关心速度，但是后者要关心的多得多。

投资者策略

*执行委托单*是一门艺术，它指的是投资者使用某种或者多种策略的组合，以最为优惠的条件执行某笔想要达成的交易。使用软件自动执行委托单的投资者有时称为**算法交易者**。其实这个概念并不是十分准确，不适合按照其字面意义来理解，因为算法是一个笼统的说法，凡是为完成一些任务而良好定义的指令集合都可以称为算法。启动汽车的步骤是一个算法，调制玛格丽塔酒的秘方是一个算法，并且做市商所做的事情也是算法——但是，一般并不把他们称为算法交易者。想想看吧。因为使用自动交易技术的投资者称为**算法交易者**，他们使用的策略也就称为**算法策略**。有时候也称为**自动交易策略**、**黑箱交易策略**，或者**机器人交易策略**（这些称谓也能很好地描述做市商所做的事情，不过这个就先别管了）。

市价委托单

投资者想要交易的时候，最简单的方式就是提交一份市价委托

单，这份委托单将以当前市场报出的申买价卖出，或者以申卖价买入（参见图 3-1，其中的圆圈、方块、三角和数字表示不同的交易者以及他们委托单的数量）。这样做的坏处就是需要支付买卖价差，而我们知道，投资者一般是不愿意支付买卖价差的。但也有些投资者是愿意掏这个钱的，比如，有时他们很想从投资组合中去掉某只股票，这时他们就不管价格多少，市价平仓出场了。

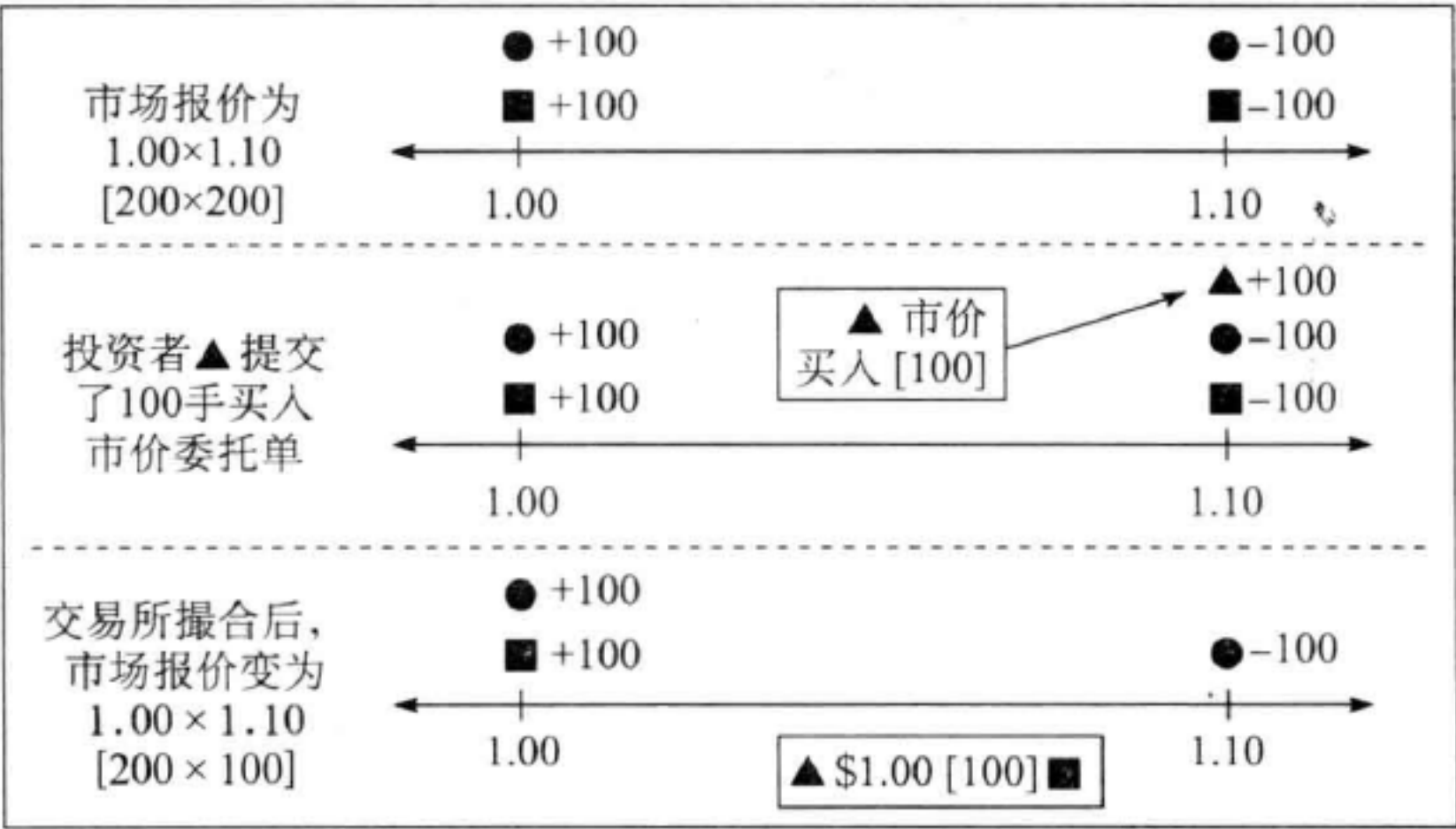


图 3-1 投资者策略：市价委托单

如果是一个市价买单，投资者表达了支付当前最优申卖价的意愿，并且期待能以这个价格成交。如果是市价卖单，他则愿意以最优申买价卖出，并且期待能够如愿。我说“期待”，是因为其委托单不一定能够及时达到并以市价成交。有可能有其他委托单抢在他前面，把申买或者申卖委托给吃掉了，或者是他的委托单还没有到的时候，市场的申买和申卖委托又修改过了。因此，即使在这么简单的情况下，执行速度也能带来很大的差别，尤其是投资者认为市场趋势正朝对他不利的方向运动并且将持续下去的时候更是如此。当然，也有可能市价委托单成交之前市场价格发生了对其有利的变动。

在结束市价委托单的讨论之前，你可以想一想，要是有一个投资者

提交了一份巨额委托单，比如买入 10 万份，但是全国最优买卖报价上面只有 1 000 份申卖单，那么会发生什么？很有可能交易所会让这个委托单“走过委托单簿”，用越来越高价位上的申卖单与之成交，直至成交结束。这个委托单会打穿好几层申卖委托，让那些被动的流动性提供者得到一个非常有吸引力的成交价格（不过对买方而言就不那么有吸引力了）。这些提供者接下来会乐不可支地把他们的买入报价提高一点，以便能完成反向交易。总而言之，一个大额买入委托单会推高市场的买卖中间价，这也印证了供需法则的作用。

在其他的一些投资者策略中，投资者会提交限价委托单，或者是指定了成交价格的委托单，这些情况我们在上一章已讨论过。

讨价还价

在接下来讨论做市商策略时，我们会看到，流动性提供者常常报出的并不是他们的最优买价和最优卖价，而是次优报价，甚至是次次优报价。可能一个市场上买卖双方都是做市商的报价，比如 1.00×1.10 ，但是他实际愿意买入的价格是 1.01，而实际愿意卖出的价格是 1.09。

投资者可以利用这种常见现象，这样比起支付买卖价差来，最少能够得到一个略微好一点的结果。比如，要在一个 1.00×1.10 的市场上买入，他可以提交一张 1.09 美元的买入委托（见图 3-2）。^① 如果实际上有人愿意以这个价格交易足够的数量，只是没有在委托单簿上显露出来的话，他就可以比提交市价委托节省一分钱。更有甚者，他还可以报入更低的委托价格，比如，报 1.05 美元来看看行不行（这也并非不可能，因为很多时候做市商为了完成反向交易，愿

① 如果想探测储备委托单（后面将会进行介绍），这里可以提交 IOC 委托单，或者叫不成即撤委托单。

意“以公允价值成交”或者以大约中间价的价位成交)。如果这张单子成交不了，他可以一次提高一分钱，按 1.06、1.07 美元的顺序再次报出委托单，直到买入成功。对他而言，最差的情况也就是拿到了市场申卖价，不过这也不会比提交市价委托单更差。当然，须假设在讨价还价的时候市场申卖价没有发生对他不利的变化。

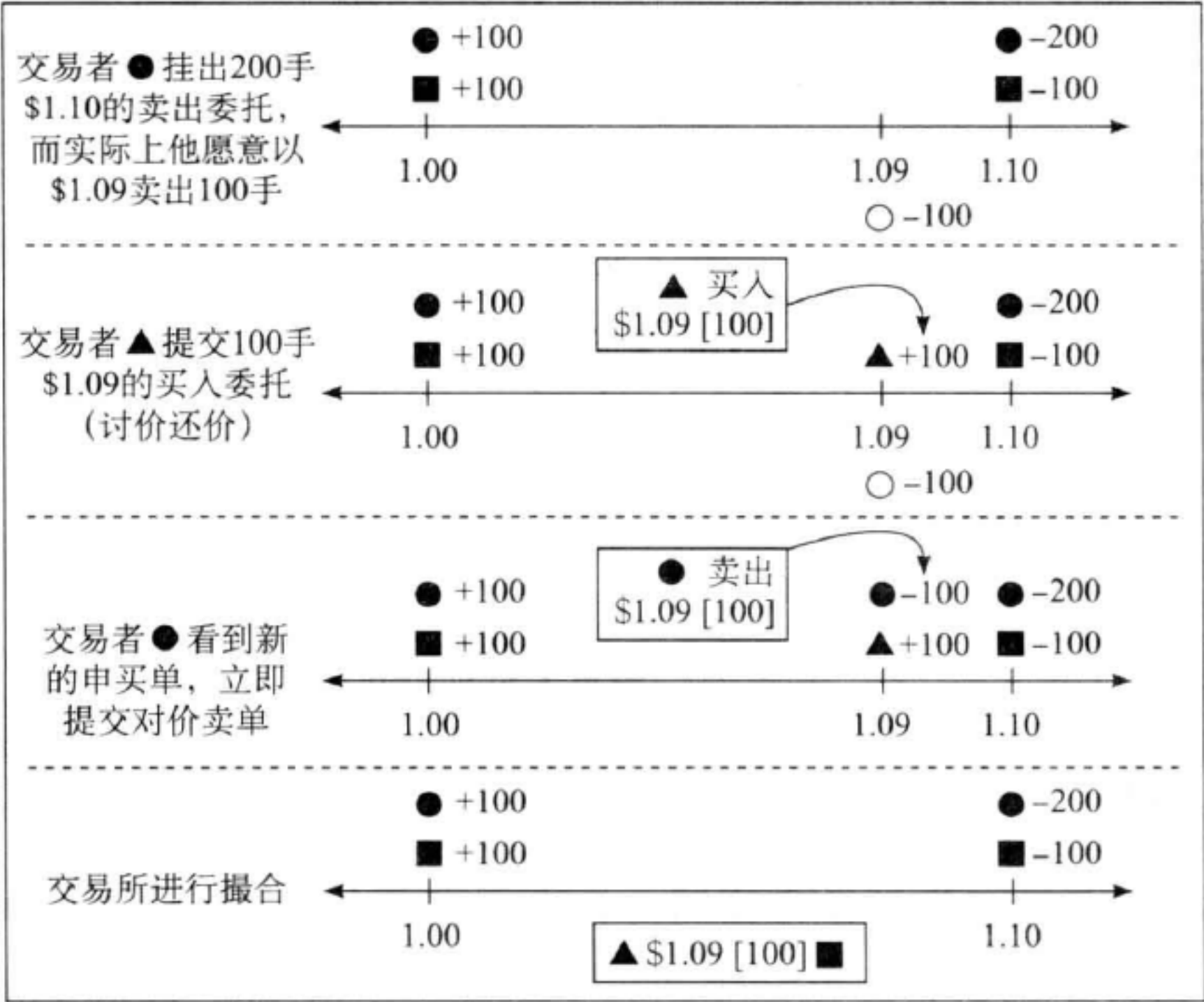


图 3-2 投资者策略：讨价还价

加入做市商

理想的情况下，投资者当然愿意以市场申买价买入，以申卖价卖出。也没什么理由他不可以提交一张限价委托单以当前最优买报价申买，或者是以最优卖报价申卖（见图 3-3）。此时，他就加入了做市商的队伍，并且采用被动的方式进行交易。如果有可以配对的委托单到来，他就可以获得成交。当然，如果交易所采用了价格－时间优先的机制，他就得排队等待，在决定是否使用这个策略时，也

得把此因素考虑在内。

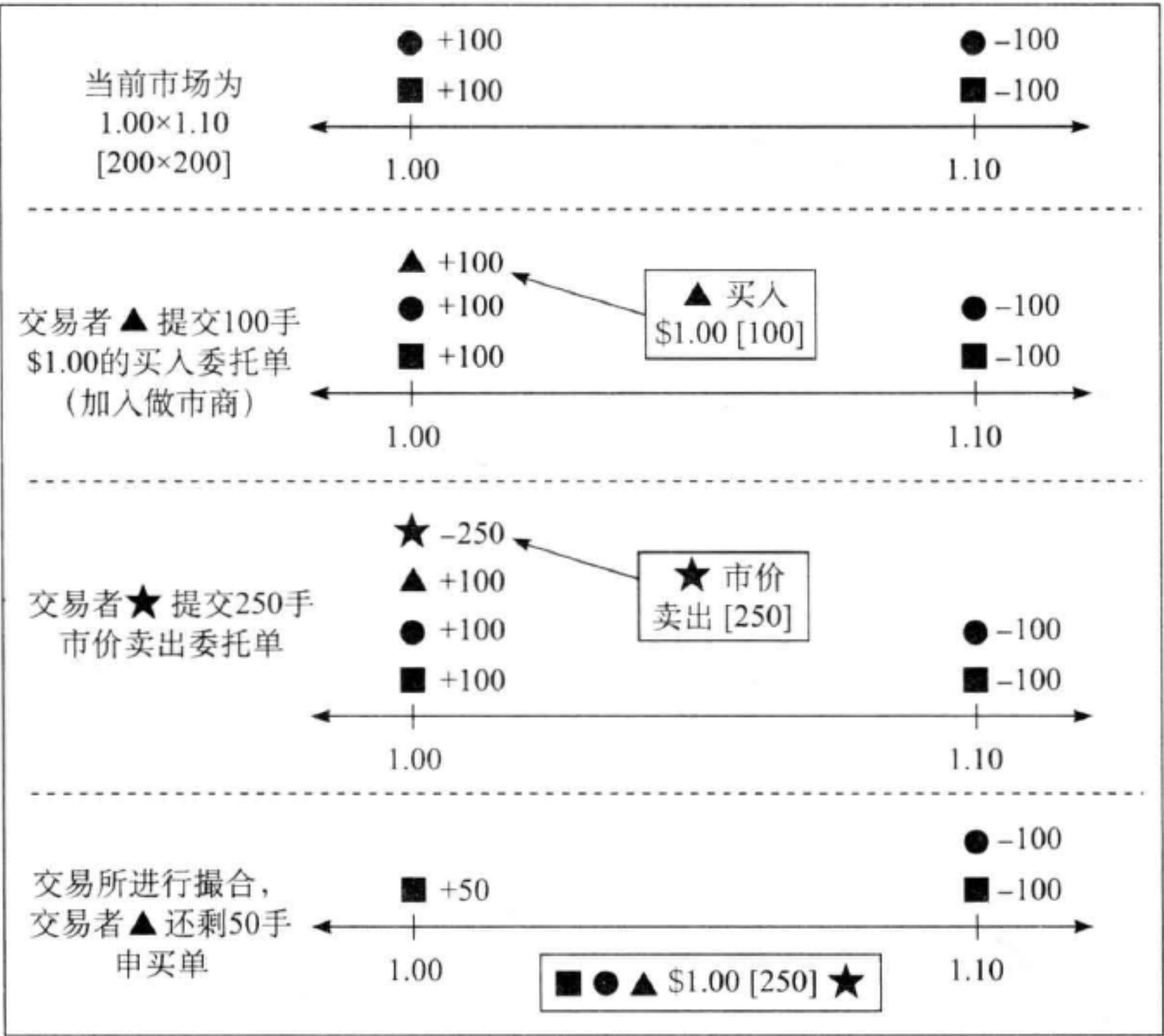


图 3-3 投资者策略：加入做市商

在决定是否加入做市商时，一个很重要的考虑因素就是市场申买和申卖委托单数量的平衡性。如果一个市场处于 250|1.00 x 1.10|5 000（为了说明情况，我们故意举了一个极不对称市场的例子）的状态，而你想要卖出，这个时候你可能就不太想自找麻烦去加入申卖的队伍，因为与你竞争买单的流动性提供者实在是太多了。但如果你想在这个市场中买入，按照这个思路，你就会很愿意加入申买者的队伍，因为此时买单少对你是有利的。另外，你还得考虑你的委托单的市场影响（例如，有可能你的加入就使得市场朝你不利的方向运动）。在后文讲到做市商策略的时候，我们会对这个现象进行更详细的讨论。

顺便说一下，提供 VWAP 价格保证的经纪自营商基本上一定会采用加入做市商策略。VWAP（念做“vee-wop”）是交易量加权平均价的简写，你可以很粗略地把它看成是某个时间段之内买卖中间价的平均值（一个与之相关的概念是 TWAP，或者叫时间加权平均价）。例如，在我们 1.00×1.10 的市场上，经纪商可能会向客户保证获得 1.05 或 1.06 美元或者其他某个中间价格，这是因为他们有信心用类似于投资者策略的方法交易获得这个价格。

下面一个策略要处理的对机构投资者而言显然是最头疼的问题，即如何最小化其委托单的市场冲击。在这个策略的每一种情况之下，投资者的委托单都在委托单簿中等待成交，但是对于其他的市场参与者，他们的委托单是部分或者全部不可见的。

储备委托单

有些市场允许把委托单放在委托单簿里，但是不显示出来。例如，市场显示的报价是 $700|0.99 \ 500|1.00 \times 1.10|500 \ 1.11|900$ ，但实际上 1 美元的申买价不止 500 手委托单，而是有 900 手。也就是说，还有 400 手订单储备在那里（见图 3-4）。这时如果有一张 700 手的市价卖单到达，这 700 手委托单将全部在 1.00 美元的价位成交。如果最优申买价上没有储备委托单的话，那么这个市价委托单将有 500 手在 1.00 美元的价位成交，200 手在 0.99 美元的价位成交。

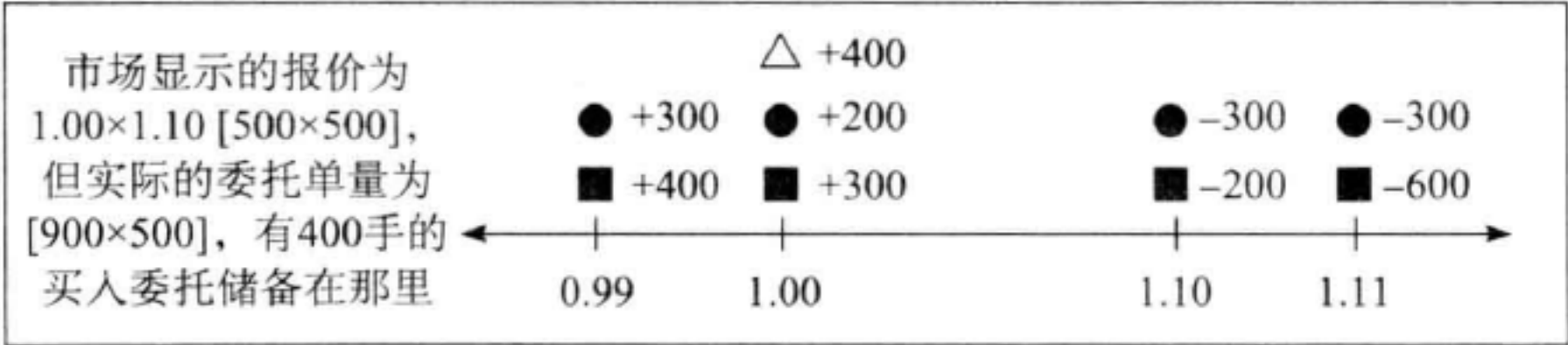


图 3-4 投资者策略：储备委托单

冰山单

储备委托单的一个变种是冰山单（iceberg order），也叫做隐藏数量委托单。在这种情况下，交易所只显示一部分的委托单，并让剩余的委托单处于储备状态。例如，市场显示的状态是 1 200|0.99 600|1.00 × 1.10|500 1.11|900，但有可能实际上 1 美元申买价那里不止 600 手委托单，而是有 3 500 手，其中 2 900 手被隐藏起来了（见图 3-5）。也就是说，有一张报价为 1 美元的 3 000 手买入委托，但是投资者知会交易所每次只显示 100 手。这 100 手被吃掉之后，交易所会再补充显示 100 手，直到所有 3 000 手委托单全部成交。另外，如果有人以多于 600 手的委托单来和申买成交，交易所一般会用储备委托单（这里是 2 900 手隐藏委托单）与之成交，直到储备委托单消耗完毕，而不是拿下一档的价格（这里是 0.99 美元）与之成交。

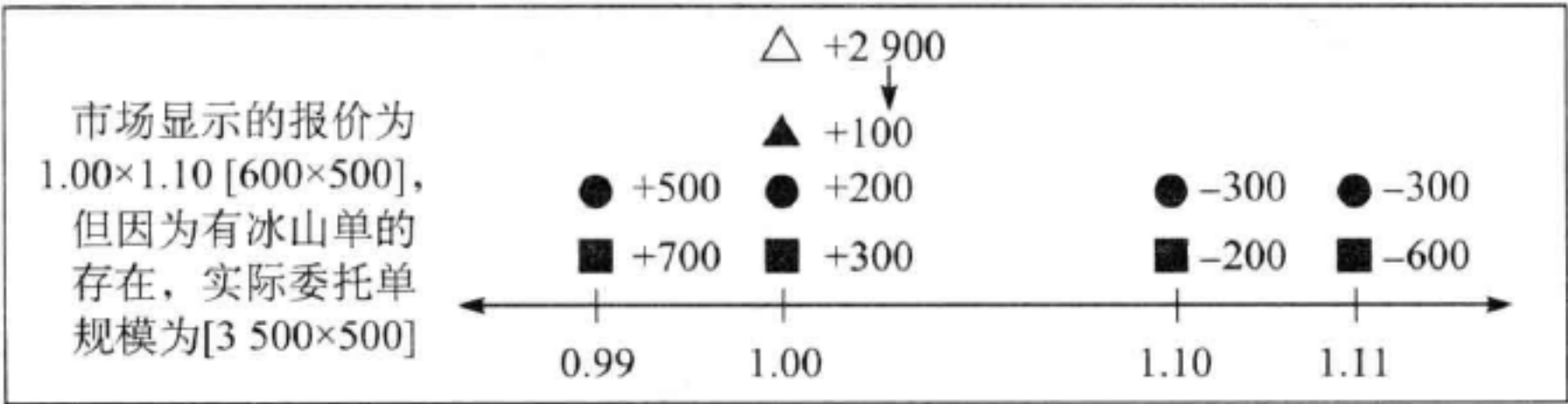


图 3-5 投资者策略：冰山单

时间分割

时间分割与冰山单的想法类似，即每次只显示委托单的一部分，这样，一笔大规模委托单看起来就像是很多“块”彼此不相关的小委托单。对于一个典型的冰山单而言，只要显示的那块成交，新的一块就会显示出来。而对于时间分割委托单来说，需要显示的那一块成交，并且再过一段时间之后（比如两秒钟。另外，更为精细的时间分割方法会采用随机时间间隔，以避免被其他人探测出来），新

的一块才会显示出来。同样，这个步骤会重复下去，直到所有委托单全部成交或者剩余委托单被撤销。

做市商策略

我们知道，纯粹的做市商没有任何动力持有证券，他们进行交易仅仅是为了赚取买卖价差（或者是至少赚到一份返还，这一点我们在后面可以看到）。在理想的情况下，他会每天都同时以申买价买入，以申卖价卖出，然后晚上高高兴兴地下班回家。但实际上，他的工作可比这辛苦得多。

在另一边等待

对于投资者而言，最基本的策略是市价单策略，对做市商来说，最基本的策略是一旦有人和他成交，他就在市场的另一边被动等待其他人和他成交。如果有人打掉了他的申买单，迫使他们买入，他们就会等待有人再打掉他的申卖单，使得他们能够卖出。或者是有人打掉他的申卖单，迫使他们卖出，他就会等待有人再打掉他的申买单，使得他们能够买入。值得注意的是，卖出一份证券时，并不需要先拥有此证券，可以通过卖空来进行申卖。一般而言，进行卖空时所需要遵循的规则和其他人并无二致，找到一个借出方，标明他的申卖属于卖空交易。总而言之，即使他手头没有仓位，并且一天之中的第一笔交易是有人打掉了他的申卖单，其实也并没有什么关系。

我们已经知道，这种策略的风险很大，因此，很少有做市商会这样完全被动等待。如果在他等的时候，市场发生了不利于他的大幅运动，那可就不好受了。如果他买入之后市场就下跌了，他就很可

能需要以更低的价格卖出。如果卖出之后价格上涨了，他就很可能需要以更高的价格买入。

倾斜市场

这里的想法是，如果有人和你交易，你会在市场的另外一侧改善报价，以增加反向成交的概率，并且在成交的那一侧降低报价，以减小增加风险敞口的概率。如果有人打掉了你的申买单，你会改进你的申卖价并且降低申买价。如果有人打掉了你的申卖单，你会改进你的申买价并且提高申卖价。按这种方式不对称地在市场上报价称为**倾斜市场**，即根据你目前的头寸和你想要达到的目标来进行报价。做市商的头寸又称**存货**，这种倾斜报价方式也是**存货风险管理**的一个例子。我们在之前提到过，做市商很难做到完全配对，一笔交易和另一笔交易完美抵消。更有可能的情况是做市商会在市场的一边或者另外一边累积头寸。当存货增长时，做市商很可能会相应地增大倾斜程度，以平抑自己的单边头寸。

刮取返还

如果做市商知道只要交易 1 000 股，他就可以得到一些返还（比如说 1 美元）[⊖]，并且市场买卖价差非常窄的话，他可能会进行一种称之为**刮擦**（scratch）的交易，平进平出完成一次双向交易。举个例子，假设当前市场报价是 1 000|1.00 × 1.01|1 000，并且所有申卖委托单都是做市商报出的，此时有人打掉了他的卖出委托单。那么做市商就可以把他的申买价提高到 1.01 美元，使得市场变成比如说 1 000|1.01 × 1.02|1 000，并且等待有人吃掉他的申买价（这也是他

⊖ 截止到 2009 年 3 月，纽约证券交易所做市商返还的标准是每股 0.001 美元，也就是这里例子中的每千股 1 美元。交易量大的做市商得到的返还会更多一些。

刚刚卖出的价位)。刚刚的卖出和随后的买入都可以为他带来一定的返还。如果每次 1 000 股委托都能完全成交，他就可以赚到 2 美元，虽然他买入和卖出的价格是完全一样的。

隐藏最优报价

做市商没有必要亮出自己愿意交易的最优价格。你可以报出比你的最优买价和最优卖价略微差一点的申买和申卖，然后观察市场上是否有人报出更优价位，并且主动吃掉这些委托单。如果你看到这些委托单的价位符合你愿意交易的价位，那么在这种情况下——也只有这种情况下——你就可以提交一张委托单与之成交。例如，你在一个状态为 $500|1.00 \times 1.10|500$ 的市场上提供双边报价，但是实际愿意买入的价格是 1.01 美元，实际愿意卖出的价格是 1.09 美元。如果限价委托单簿上出现了一张 1.01 美元的卖单（可能来自于一个讨价还价的投资者），你会立即递交一张 1.01 美元的买入委托单与之成交。这里的想法就是在被动交易的同时，如果遇到有利的市场情况，也可以进行主动交易。

那为什么不直接在市场上挂出 1.01×1.09 的报价呢？如果你报出的买卖价差较宽，在一个市价卖单到来并且你处于最优买卖报价的情况下，你报出 1.00 美元的买价显然比报出 1.01 美元要好。换句话说，你可以让可见的买卖价差大一点，这样和市价单成交时更有利，并且在遇到比市场价更好的限价委托单时，你还可以利用你隐藏的、更紧的报价来与之成交。

吃掉动作慢的

这个策略说明了为什么在其他条件都相同的情况下，被动交易比主动交易的风险要大得多。这里的场景与上一个策略完全相同，只

56 打开高频交易的黑箱

All About High-Frequency Trading

不过市场价格变动十分迅速，并且一些做市商的反应比其他做市商要快一点——实际情况基本上必然如此。和之前一样，你的被动市场报价总是比你的主动市场价格要宽一点。如果你的主动价格优于被动报价一分钱，那么当你的被动报价变动时，你的主动市场价格也会随之改变。到这里你所要做的事情也就差不多了。现在，市场发生了快速变动，改变报价以适应新的市场时，你可能会发现有些人的被动委托单还没有来得及撤销，并且与你愿意主动成交的价格相遇了。

例如，你和其他一些被动交易者一起，在一个 1.00×1.03 的市场上提供双边报价，而实际上你愿意主动交易的价格是 1.01×1.02 （见图 3-6）。如果市场价格突然上涨，你把被动市场报价改为 1.02×1.05 （其他大部分被动交易者也会如此，但并不是所有交易者都如此），你的主动市场价格变为 1.03×1.04 ，但是这里还有一个比你们都慢的被动交易者，他还在 1.03 美元处申卖。你的主动市场委托单就可以与这个卖单成交了。

注意，更紧的主动市场价格并不是这里的必要条件。你还可以用你的新的被动市场报价与动作慢的被动交易者成交。只不过是在更紧的市场价格的情况下，较小的市场变动也能带来这类交易机会。

跳一分钱

做市商手头累积了净头寸的时候，他总是面临着反向交易无法盈利的风险。更糟糕的是，如果他还没来得及平掉仓位，价格出现了对他不利地快速变动，那他的损失就变得很难控制了。由于这种风险的存在，他在报价的时候会比原本应有的更加小心谨慎。为了降低风险，以让自己报出的价格更加激进一点，做市商可以观察最优买卖报价上是否有哪一边出现了规模相对很大的委托单。如果做市

商可以很快改变报价的话，他就可以把这些委托单当做潜在损失的支撑垫，或者是一层安全网。

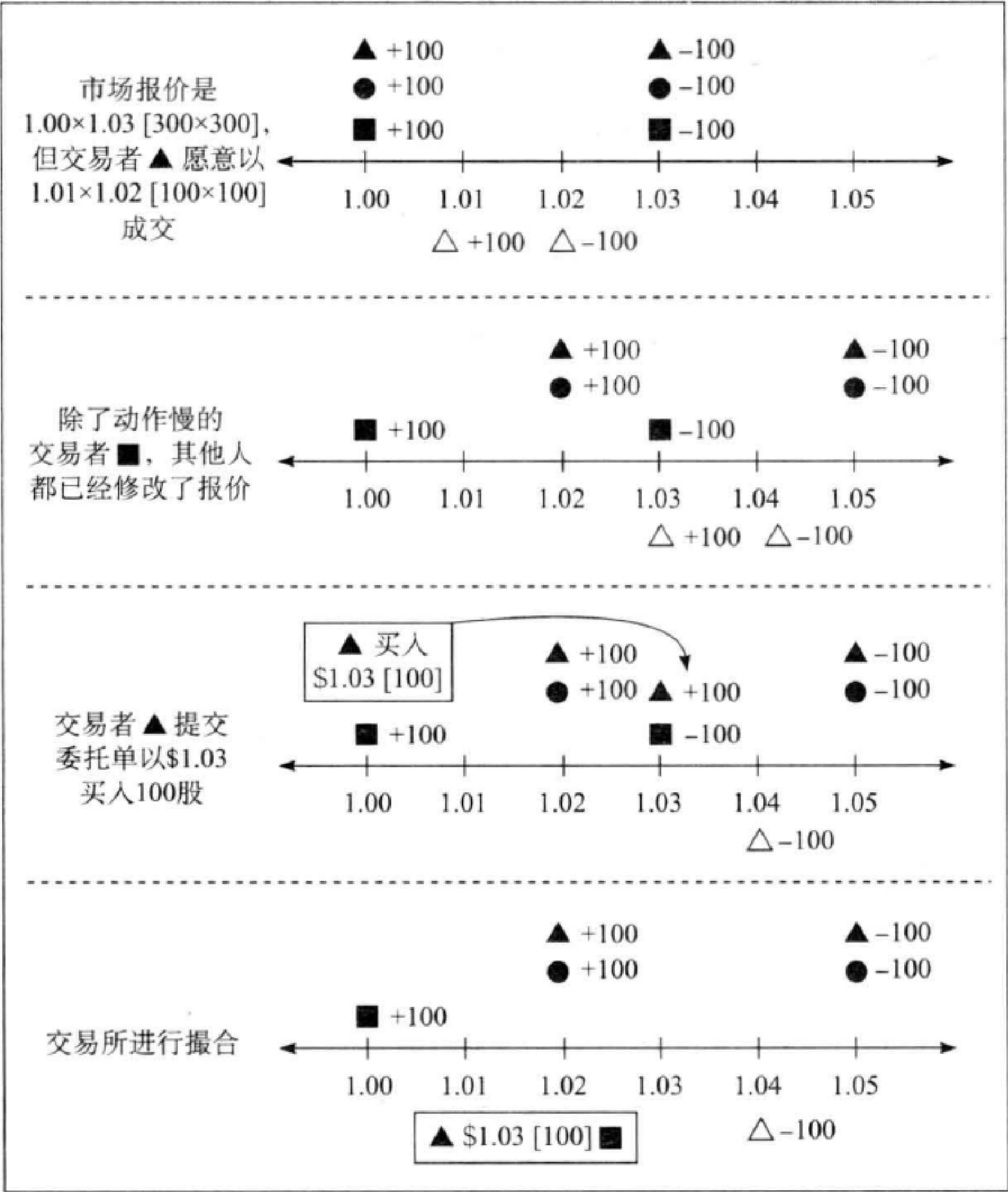


图 3-6 做市商策略：吃掉动作慢的

例如，当前的市场状态是 200|1.01×1.03|200，接着变成了 3 200|1.01×1.03|200（见图 3-7）。你知道有人想要以 1.01 美元的价格买入 3 000 股（现在还不考虑你把这份委托单当做一头大象时的做法，这是接下来要讨论的话题）。得知此信息后，你可以把申买价提高一分钱到 1.02 美元（这就是跳一分钱的意思），并且有人吃掉了你

的申买单。你得注意，不要买入超过 3 000 股。如果市场价格上涨到 1.03×1.05 ，你就可以通过吃掉 1.03 美元的申买单完成反向交易（假设申买量足够的话），并且每股盈利 0.01 美元。你还可以在平掉仓位之前先等一等，看看市场价格能不能涨得更高，让你赚得更多一点。

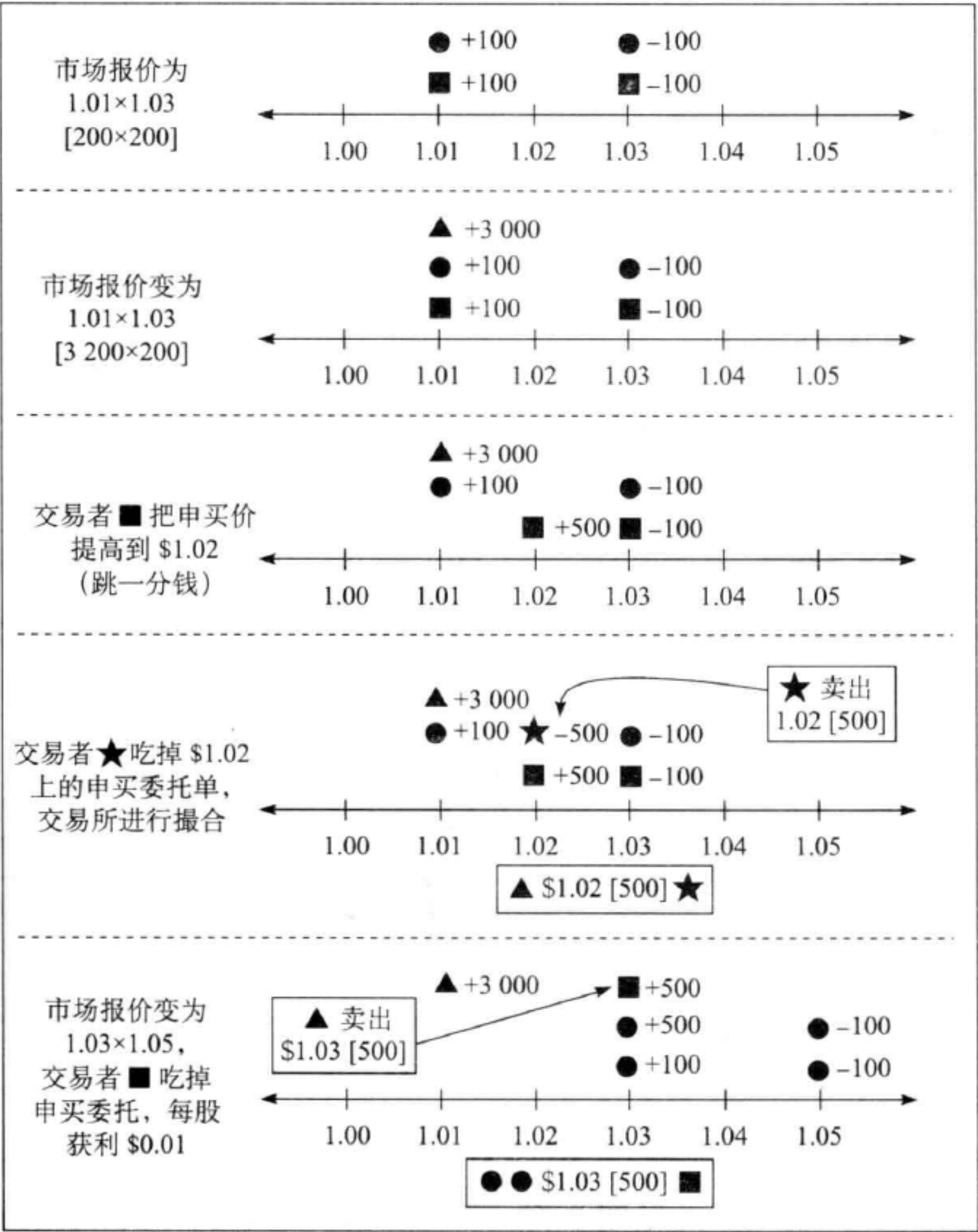


图 3-7 做市商策略：跳一分钱，价格上涨

话说回来，如果过了一段时间价格还没有上涨，你宁愿截断亏损也不愿意再等下去，你就可以把你的头寸卖给那个报价为 1.01 美元

的买家，他的委托单会把你的亏损限制在一分钱（见图3-8）^①。当然，这需要假设在你想要平仓的时候，3 000股申买委托单还在那里。因此，在这种情况下，非常高的执行速度是至关重要的，因为你需要在为你垫底的申买者有机会修改或者撤销其委托单之前完成所有的动作。虽然这个策略称为跳一分钱，但在最小变动价位小于一分钱的市场上，在使用这个策略的时候显然不需要一定要跳动整整一分钱。实际上，你只需要提高一个最小变动价位就行了。比如说，如果最小变动价位是0.001美元，在这个例子中跳一分钱交易者只需要把申买价提高0.001美元。

推动大象

在前面的章节中，我们已经看到自然的趋势就是需求的增加会推动价格上升，反过来供给的增加则会导致价格下降。这里的策略就符合这个原理。比如，市场处于500|1.00×1.01|500的状态，接着突然变成5 500|1.00×1.01|500。这个新加入的大块头（大象）显然胃口不小。他想买入5 000份证券，但不想支付买卖价差，因此他加入了申买者的行列，而不是直接使用市价指令。如果你作为一个做市商，假定这只大象实际愿意支付的价格不止1.00美元（这个假设并非不切实际）的话，你就可以买掉1.01美元上的所有卖单，并且把申买价提高到1.01美元（见图3-9）。现在市场变为1.01×1.02。如果大象仍然想买的话，他就得提高买报价。如果他真的这么做了，你可以故技重施：买掉1.02美元上的卖单，并把申买价提升到1.02美元，从而推着大象又得提高他的买报价了。

这个过程可以重复多次。你期待有朝一日大象提交的买入报价会

^① 因为这个策略的收益向上无限，而向下有限，因此其收益与看涨期权有点类似。

高于你买入时所支付的平均价格，比如说是 1.05 美元，这样你就可以吃掉这些申买单，把你推高价格时一路买到的证券全部卖给大象，并且赚到一笔利润。不过这个策略并非万无一失，因为大象随时可能撤销他的委托单，那样你可能就得以亏损的价格卖掉你的一部分或者全部股票了。如果大象真地想买，并且愿意跟着提高申买报价的话，你这么做就是可行的。

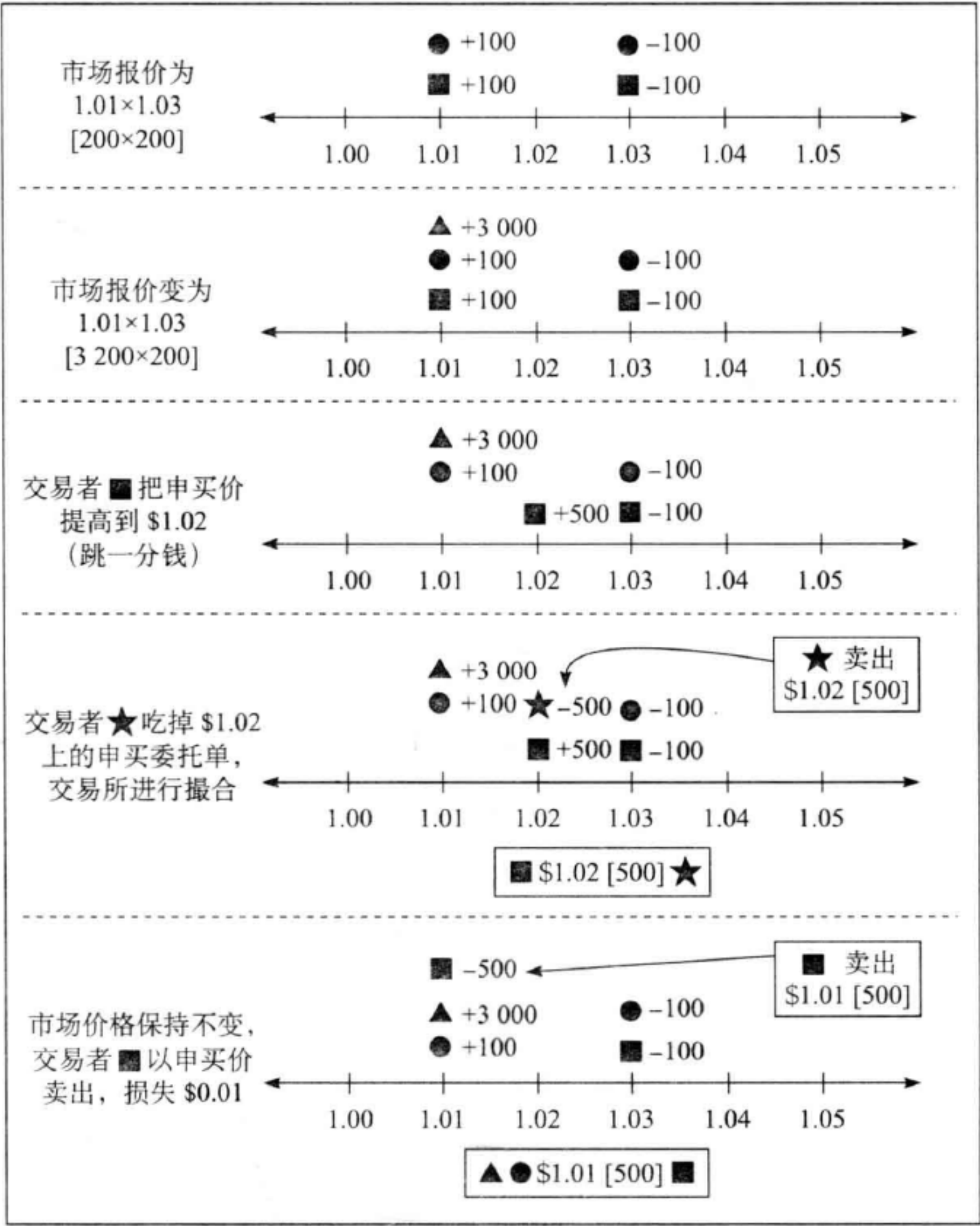


图 3-8 做市商策略：跳一分钱，价格没有上涨

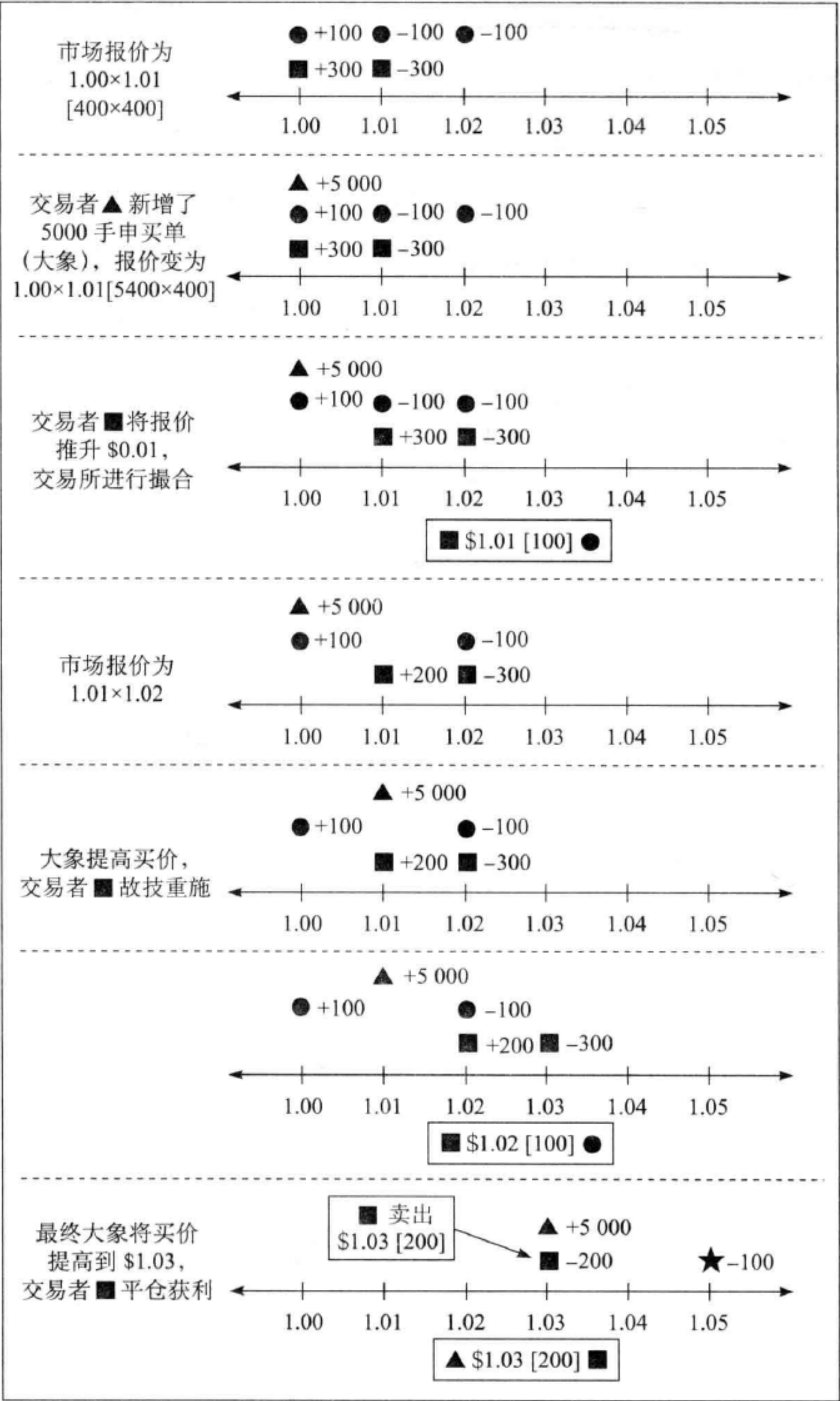


图 3-9 做市商策略：推动大象

62 打开高频交易的黑箱

All About High-Frequency Trading

我们使用了一个大额委托单来描述这个策略，但并没有明确定义多大规模的委托单算是一只大象，多大规模不算是。理论上讲，即使只有一股[⊖]，也会对市场产生买卖压力，但实际上这个影响是微乎其微的。那么要多大的委托单才能触发这个策略呢？当然，这取决于市场所显示的委托单规模的大小。在已经有上万股委托单的市场上增加 1 000 股，和在只有 100 股的市场上增加 1 000 股，其所造成的影响显然是不同的。

不过大象有多大可能会改变其价格？能够改变多少呢？要算出来多大的委托单才能移动市场，能够移动多大幅度，以及达到期望移动幅度的概率，这些都是开发这样一个策略所需要进行的复杂数据分析的很好例子。策略的开发过程可能需要分析如山的历史数据，进行大量的测试和反复的实验。然而，随着时间的流逝，这些精明的做市商们可能会发现一个相当实用的模型，并且通过这个模型迅速地给出这些问题的答案。

算不算是抢先交易

推动大象与非法的抢先交易看起来有些类似。抢先交易者根据新出现的尚未成交的委托单信息更改自己的交易行为，通过牺牲这些委托单所有者的利益来为自己谋利。乍看起来，这种行为与抢先交易并无二致。但是这种行为的辩护者会马上就指出，信息这个词应该改成非公开信息。他们会声明（或者通过他们的律师，用比这里法律得多的语言声明），新委托单的出现是一个公开信息，而利用公开信息来更改自己的交易行为本身是公平的。在这里，市场中出现大象是所有人都知道的。每一个市场参与者都有机会对此做出反应——也可以说是

⊖ 只买入或者卖出一股股票的委托单，或者是只买卖一张期权或者期货合约的委托单。

“抢在这个委托单前面”——他们这么做是符合自身利益的，并且也是符合天然的供需规律的。至于非法的抢先交易，辩护者可能会继续说道，它指的是通过一个新委托单出现的非公开信息来改变自身交易的行为。

举一个明显的非法抢先交易的例子。比如某市场的报价是 $500|1.00 \times 1.03|500$ （为了表述简单起见，我们假设只有一家交易所）。一个经纪商接到客户电话，要他下达一份市价委托单买入 500 股。在提交委托单之前，经纪商给一位老兄打电话，告诉马上就会下一张 500 股的市价买入委托单。这位老兄随后卖掉所有 1.03 美元价位上的申卖单，并且在 1.04 美元的价位上挂出新的申卖单。这时市价委托单到来，他就把自己刚刚以 1.03 美元价格买过来的 500 股以 1.04 美元的价格卖给了这位毫不知情的客户。由于除了经纪商之外，没有其他任何人知道客户的下单意向，因此这里的抢先交易显然是不干净的。

下面这个策略初看起来也像是抢先交易，但即使是这个策略，做市商也用的是公开信息来推测隐藏委托单是否存在。来看看吧，你是不是同意这个说法呢？

牵引冰山

前面我们已经知道，投资者可以采用冰山单的策略以避免成为大象，他可以把委托单拆分成小块分开显示，而不是暴露其全部交易意向。然而，即便如此，那些精明的做市商仍然可以找到一些冰山单的蛛丝马迹，并且像对待大象那样来对待冰山单。

比如市场报价为 $500|1.02 \times 1.07|300$ 。一个买家打掉了 1.07 美元的 300 股申卖单。但是申卖立即刷新，市场仍然保持不变。片刻以后，这种事情又发生了：有买家打掉了申卖，但是申卖立即刷新。

做市商注意到这个现象（通过观察数据馈送），并且假设 1.07 美元价位上显示的申卖单只是冰山一角，这个价位上还有大量的储备委托单。他根据过去的观测数据，推测认为水面之下至少还有 2 000 股的规模。现在 he 就把对付大象的那一套搬到这里来（不过这次是在市场的另一侧进行交易，因为这个冰山单是个卖单）。他会在市场下跌的过程之中不断卖出股票——实际上是他的卖出导致价格下跌——并且在价格更低的时候把股票再买回来（平掉空头头寸）。

首先，他会吃掉 1.02 美元上的申买单（从市场移除），并把申卖价改为 1.06 美元，这比冰山单的价位好一分钱。现在市场报价是 1.01×1.06 。如果冰山单把卖价改为 1.05 美元，做市商就会如法炮制：吃掉 1.01 美元上的申买单，并把申卖价改为 1.04 美元。如此继续下去。直到某个价位，比如是 0.96×1.00 ，这时冰山单可能已经成交并且市场降到了最低点。做市商就打掉 1.00 美元上的卖单，平掉他一路累积下来的空头头寸，用 1 美元的价格把他刚刚以 1.01 美元和 1.02 美元卖出的股票买回来。

不公平？不道义？知道有大单之后就去操纵市场？可能看起来如此，但和推动大象不同的是，这里的做市商并不确切地知道是不是真的有一张冰山单。他只是猜测——虽然他可以根据对数据的分析，猜中的概率非常高——但他还是有可能出错。如果他真的错了，他就会发现自己手头上拿了一大把亏损的委托单（比如，这些空头头寸没法平掉获利）。

掠夺性探测

在上面的场景中，做市商利用公开信息——被吃掉的申卖委托单总是立即刷新——推断市场中存在冰山单。先假设我们都认为这一点是公平的。那么做市商把卖价降到 1.06 美元，挡在 1.07 美元的卖

家前面，不让他有机会以此价位成交，这还是不是公平的呢？这就很难说了。

做市商还可以更进一步，向市场发送大量的 IOC 委托单来探测卖家能够接受的最低价位是多少，比如说，他可以在不同的价位上——一个是 1.06 美元，一个是 1.05 美元，还有一个是 1.04 美元发送 IOC 委托单，看看实际会发生什么，如果他这么做的话，是不是公平的呢？如果 1.06 美元和 1.05 美元的委托单成交了，但是 1.04 美元的撤销了，他可能就可以确切地知道卖家最低能够接受的价格是 1.05 美元，有了这个知识，他可以在下一步加以利用。有人会说（也相当有说服力），这样向市场中发送探测性的委托单——即使只是为了探测其他人的价格底线而没有其他任何目的——已经越过了创新和掠夺之间的界限。如果掠夺者比猎物的动作更快，整个事情会显得更加恶劣。比如说，那些掠夺者的装备比 1.05 美元价位上大单的卖家更好，导致卖家无法在他的委托单成交之前修改或者撤销委托单。

当前（2010 年年初），像推动大象、牵引冰山以及 IOC 流动性探测等方法都还算是符合现有监管规则的，或者最多算是一个高频交易商从中渔利的灰色地带。以后还会不会是这样，现在还不太好说。

跳过 delta

期权做市商与股票做市商的基本目标是一样的，即以较低价格买入，以较高价格卖出，并赚取买卖价差^①。但由于种种原因，比如期权合约数目众多等，期权做市商在一笔成交之后，很难通过另外一

① 期权和股票一样，也是用元和分报价，但是期权交易者常常使用隐含波动率来代指价格。因此，对期权做市有时也称为波动率套利（volatility arbitrage）或者倒卖波动率（scalping vol）。

笔期权交易（完成反向交易）来消除风险。相反，他们会通过交易其他一些经过精心选择的产品来锁定收益，或者对冲风险。比如说，他们交易了期权之后，接下来会交易一定数量的标的股票（例如交易了微软公司的期权，随后立即交易微软的股票）。如果股票仓位调整合理，股票的投资组合能够视作一个合成期权，其涨跌幅度与已交易的期权大致相等，但是方向相反。这个过程称为 *delta 对冲*，之所以叫这个名字，是因为 *delta* 作为一个灵敏性度量[⊖]，决定了期权做市商对冲时需要交易的股票数量。

标的股票的做市商可以利用这个过程之中的信息，抢在 *delta* 对冲之前进行交易。其基本想法就是观察期权市场上的大额成交，推断出做市商的持仓方向，计算对这笔期权交易进行 *delta* 对冲所需要的股票交易量，并且抢在它前面进行交易。例如，股票做市商在期权市场数据馈送中看到一笔 5 000 手 XYZ 股票看跌期权[⊖]的交易。期权的执行价格与 XYZ 股票目前的市场价格大致相等，因此，做市商知道这是一份平值期权，当然它的 *delta* 会在 50 左右。由于期权的成交价格靠近当前的期权申卖价，股票做市商可以推断出期权的做市商卖出了期权。

为了对冲卖出的看跌期权，期权做市商必须卖空股票。这里的

⊖ *delta* 表示的是期权价格如何随标的物价格的变化而变化，因此，人们用它来计算做市商需要持有多少股票来抵消掉期权的价格变动。*delta* 对冲是期权做市商众多对冲中的一种，也是至关重要的一种。还有一个敏感度希腊字母是 *vega*，它表示的是期权价格如何随着标的物波动率的变化而变化。*rho* 表示的是期权价格对利率变化的敏感度，*theta* 表示的是期权价格对时间变化的敏感度。最后还有一个字母是 *gamma*，它表示的是 *delta* 如何随着标的物的变化而变化。*vega* 和 *gamma* 只能通过交易其他期权进行对冲。*rho* 可以通过欧洲美元之类的利率期货合约对冲。在数学上做一点延伸的话，这里的每个希腊字母都是从著名的布莱克－斯科尔斯偏微分方程之中得到的偏导数，布莱克－斯科尔斯期权定价公式也是从这个方程中推导出来的。

⊖ 看跌期权赋予其买家在未来某个时刻以指定价格卖出标的证券的权利，但买家无卖出义务。看涨期权赋予其买家买入标的证券的权利。

delta 是 50，期权的交易量是 5 000 手合约，每手合约都有 100 股股票作为其标的物，因此，期权做市商必须卖出 250 000 股 XYZ 的股票（ $0.50 \times 100 \times 1\,000^{\ominus}$ ）。期权做市商一般愿意支付股票的买卖价差，因此他们通常使用市价单进行 delta 对冲。比如当前的市场状态是 25.50×25.60 ，如果股票做市商的动作比期权做市商的动作快得多的话，他就可以吃掉 25.50 美元的买价，并且改在 25.49 美元上挂出买单。期权对冲的卖出市价委托单到来的时候，它将以 25.49 美元的价格成交，而快如闪电的股票做市商就把他刚刚以 25.50 美元卖出去的股票以 25.49 美元的价格买了回来。

再谈投资者策略

做市商采用跳一分钱、牵引冰山之类的种种策略，谁知道还有没有其他什么策略，那投资者还有活路吗？答案当然是肯定的。既然做市商可以戏耍投资者，那为什么投资者就不能戏耍做市商呢？或者至少是采用一些手段让做市商不能得逞——当然，这都是在法律许可的情形之下。比如说，为了不让冰山单被牵引出来，投资者可以将委托单显示的比例和委托单更新的频率随机化，让这个单子一开始看起来就不像是一个冰山单。如果你的速度足够快的话，吓唬一下跳一分钱交易者也不是什么难事。比如，如果你真的想买的话，你可以挂出申卖单，而不是申买单，让跳一分钱交易者抢在你前面交易，然后再撤销掉委托单把他的安全网拿走。在他方寸大乱、忙于管理自己新的风险的时候，你就可以在市场的另一边买入了。

当然，没有什么理由这些技巧和技术只有卖方高频交易者能用，而共同基金、保险公司还有其他一些买方的机构投资者就不能用。

[⊖] 此处疑为 5 000。——译者注

68 打开高频交易的黑箱

All About High-Frequency Trading

例如，管道交易系统（Pipeline Trading System）公司就在尝试这么做。他们对高频交易策略了如指掌，并且开发了一些策略专门来对付它们。举个例子，他们开发了一种类似于冰山单的策略，不过只要委托单在一个交易所成交一点，他们就撤销委托单。接下来他们就把冰山剩下的部分移到另外一个市场，并且如此重复下去。这一切都让高频交易者来不及弄清楚发生了什么事以及把价格朝不利的方向推动，当然所有的过程也都必须以闪电般的速度进行。

套利策略

我们已经讲过如果同一只股票在两个不同交易所交易，形成了交叉市场时，可以进行套利，这可能是最简单的套利策略。比如 PDQ 股票在一家交易所的申卖价是 1.02 美元，而在另一家的申买价是 1.03 美元。套利者就可以简单地打掉 1.02 美元的卖单并且吃掉 1.03 美元的买单，赚到一分钱的差价。这确实太简单了，并且基本上不要任何本钱，因此，这样的机会一般消失的比闪电都要快。

不过，由于产生交叉价格的股票数量有限，并且考虑到交易成本，这种套利的总体利润是受到限制的。在我们的例子中，如果 1.02 美元的申卖有 10 000 股，1.03 美元的申买有 2 000 股，那么你最多也就能赚到 20 美元。这里也像其他有关套利的例子一样，假设交易的成本不超过套利产生的利润。如果执行一股套利的成本超过一分钱，那就没有必要进行套利交易了。

实际上，这也是实际进行套利交易的一个重大障碍，特别是对于经纪费用相对较高的散户更是如此。那些交易量大的套利交易者会进行大量的投资，以维持极低的边际交易成本，比如他们会购买直

接在多个交易所交易的权利，这样就可以绕过经纪商。你可以把你的家底都押上，来赌很多高频交易公司已经在这类事情上面投入了数百万的资金。因此，股票和 ETF（它们可以用和股票一样的方式进行套利）不论在哪个市场交易，其价格基本上总是一样的。

在权益类证券这个大超市里，套利机会并不局限于交叉市场中的单一股票。并不是只有相同证券在不同市场交易，同时但不同价时才有套利机会。只要两种证券或证券的投资组合的价格彼此关联、必须共同变动，它们之间就有套利的可能。而我们的超市中这样的证券实在是太多了。

ETF 和一篮子股票

考虑一下 Spider ETF（代码 SPDR），我们知道，在任何时刻，它的价值都是由组成标普 500 指数的一篮子股票的总体现值所决定的。想象一下，你持有这 500 只股票多头所构成的投资组合。自然而然，只要任何一只股票的价值发生了变化，你的投资组合的价值就会发生变化。股票分红时也是如此。现在，再想象一下你持有 Spider ETF 的多头头寸，它的价值同样会因为标普 500 成分股的价格变化或者分红而产生变化。

如此一来，套利者会在每一刻仔细跟踪这 500 只股票篮子的价值和分红情况，用这些数据计算 Spider 的合理价格，并与 Spider 的实际市场价格进行对比。接下来他要寻找一个交叉市场。如果他发现 Spider 的净市场申买价超过了股票篮子的净市场申卖价，他就卖出 Spider 并且买入股票篮子。如果股票篮子的净申买价超过了 Spider 的净申卖价，他就买入 Spider 并且卖出股票篮子。

市场上的 ETF 种类繁多，并且总是有新的 ETF 上市，每个新的 ETF 都带来新的套利机会。目前比较流行的宽基指数 ETF 除了标普

500 指数的 Spider 之外，还有纳斯达克 100 指数的 QQQQ，以及罗素 2000 指数的 IWM。比较受欢迎的小型指数 ETF 包括新兴市场指数的 EEM，和跟踪金融板块股票指数的 XLF。

很显然，交易速度是这个策略成败的关键。只有套利者能在价格变动之前（例如在 ETF 和股票篮子的相对价值回归之前）完成交易，套利才是可行的。你要知道，其中一笔交易可能是要交易 500 只不同的股票呢。因为不止一家公司对这类套利虎视眈眈，这就很明显，他必须得比其他人都要快才行。

实践中，很多经纪商可以让你用一张委托单交易一篮子股票，或者是方便在 ETF 和股票篮子之间转换。这种服务是收费的，当然，在估算套利是否盈利时需要把这笔费用算进去。毫无疑问，ETF 套利比单只股票套利要复杂一点。但是它还是可以完成的，就像其他各种套利也可以完成一样，并且它的完成方式是相当简单直接的。

期货与一篮子股票，期货与 ETF

因为指数期货合约的价值最终也是由指数成分股的价值所决定的，这种合约可以像 ETF 一样进行套利。实际上，期货合约和 ETF 之间也可以进行套利。表 3-1 列出了一些比较常见的指数期货、它们各自所在的交易所，以及相对应的 ETF（在多个股票市场交易）。

表 3-1 期货合约与对应的 ETF

交易所	期货	ETF
芝加哥商品交易所（CME）	标普 500 指数期货（SP） E- 迷你标普 500 指数期货（ES）	SPDR
芝加哥商品交易所（CME）	纳斯达克 100 指数期货（ND） E- 迷你纳斯达克指数期货（NQ）	QQQQ
芝加哥商品交易所（CME）	道琼斯指数期货（DJ） E- 迷你道琼斯指数期货（YM）	DIA
洲际交易所（ICE）	罗素 2000 指数迷你期货（TF）	IWM

期货与期货

有些期货合约之间可以很容易进行套利。例如，CME 上市的不仅有 ES（迷你合约），还有 SP（大合约），这两个合约的主要不同之处就在于合约乘数，SP 的乘数是 250 美元，而 ES 的乘数是 50 美元。另一个不同之处是，在交易时段之内，SP 只在公开喊价市场中交易。然而，交易时段之后，SP 可以在 CME 的电子平台 Globex 上交易。因此，可以比较容易地自动化监测 ES 和 SP 的价格，并寻找价格不一致的机会。这种套利方法也可以应用在纳斯达克 -100 指数期货的迷你合约（NQ）和大合约（ND）上面。提醒一下，大合约不像迷你合约那么活跃，但还是有交易的。如果你已经做好了交易两个合约的准备（很多高频交易公司都是这样），即使是所有人都下班回家了，你的机器也会在那里兢兢业业地寻找套利机会。

期货与期权

要是对一价定律给出精确一点的定义，就是如果两只证券和 / 或投资组合未来的现金流相等，那么它们必须具有相同的价格。按这种方式来看证券的话，我们就可以找到很多有关期权的套利策略。

首先比较一下我们讨论过的三种证券的收益，即期货、看涨期权和看跌期权。我们将用相当常规的图表来画出这几种证券的收益曲线，这些曲线将收益简单地表示成期货交割时（对期权而言是到期时）交割价（对期权而言是执行价）和标的物价格的函数。

在图 3-10 和图 3-11 中，我们可以看到，对于一个交割价格为 4 的期货合约而言，标的物价格和期货合约的收益之间呈现出一种简单的线性关系。这里我们完全不考虑交易成本和贴现因素，但在实际中，这两点都是很关键的（复习一下第 2 章中有关期货的知识可能会有所帮助）。对于期货多头而言，如果现货价格是 3，则期货收益

是-1，这表示持仓有亏损。其他价格的收益也照此计算。对于空头而言，收益则是完全反过来，这一点应该与你的预期相符。

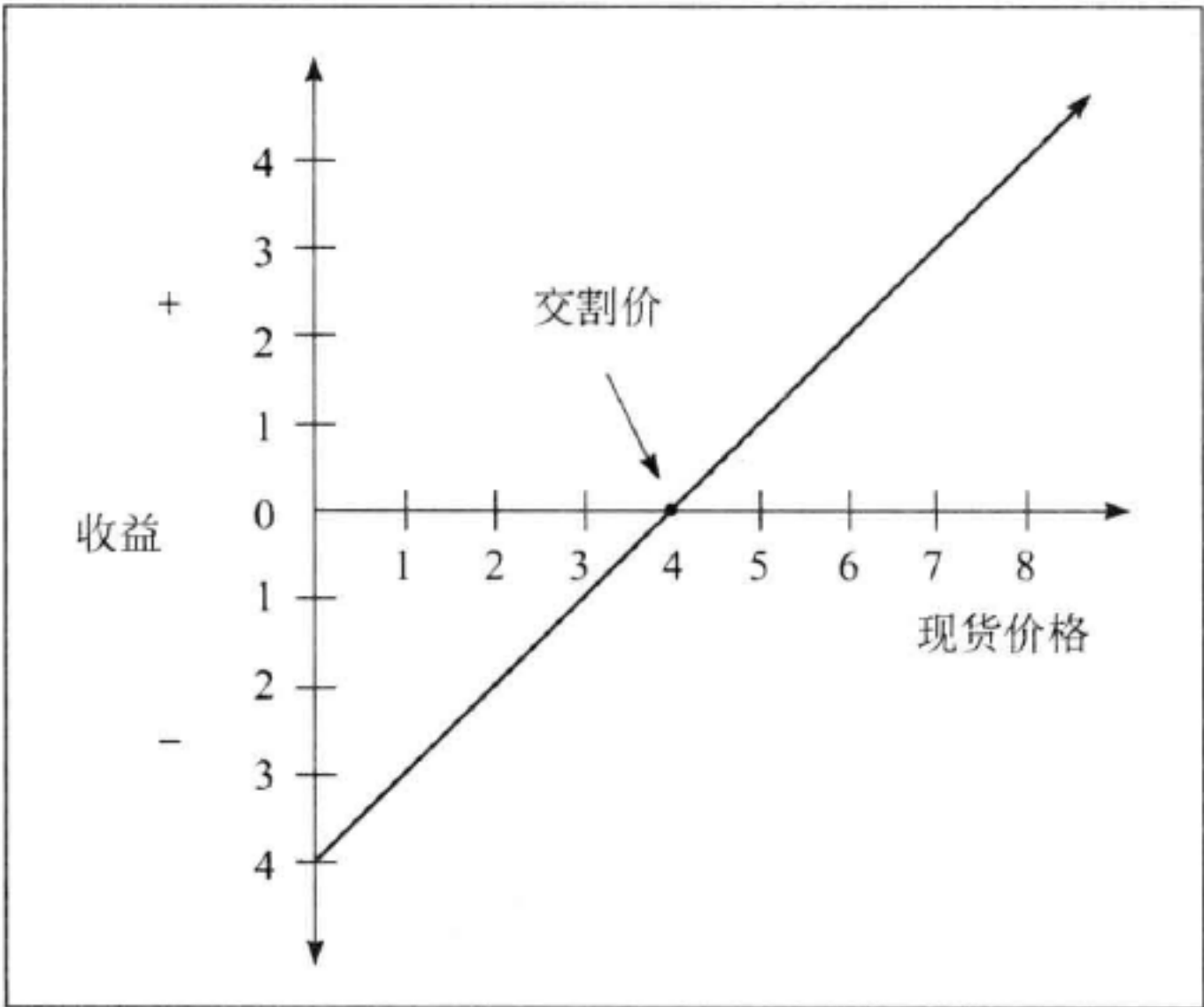


图 3-10 期货多头收益

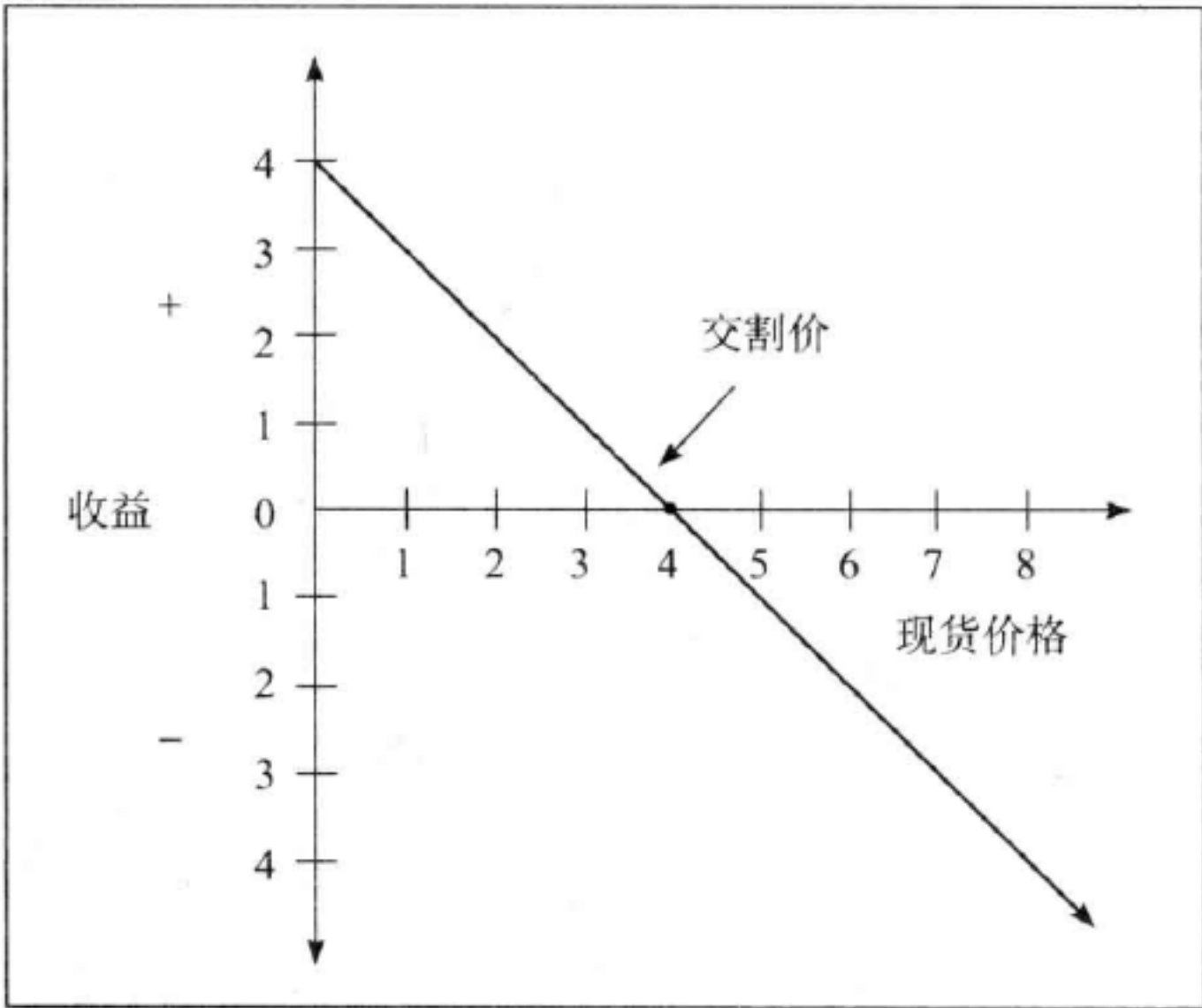


图 3-11 期货空头收益

期权的收益有一部分和期货一样也是线性的，但是对于多头而言亏损有限，对空头而言收益有限（见图 3-12）。为了简单起见，我

们不考虑期权的权利金。如果考虑进去的话，期权空头这根“曲棍球棒”似的收益曲线会上移，多头收益曲线会下移。这些期权的收益曲线与期货都差别很大。但是，如果我们把一个看涨期权的多头和一个看跌期权的空头组合在一起，形成一个所谓的合成期货合约，那情形又会怎样呢（见图 3-13）？

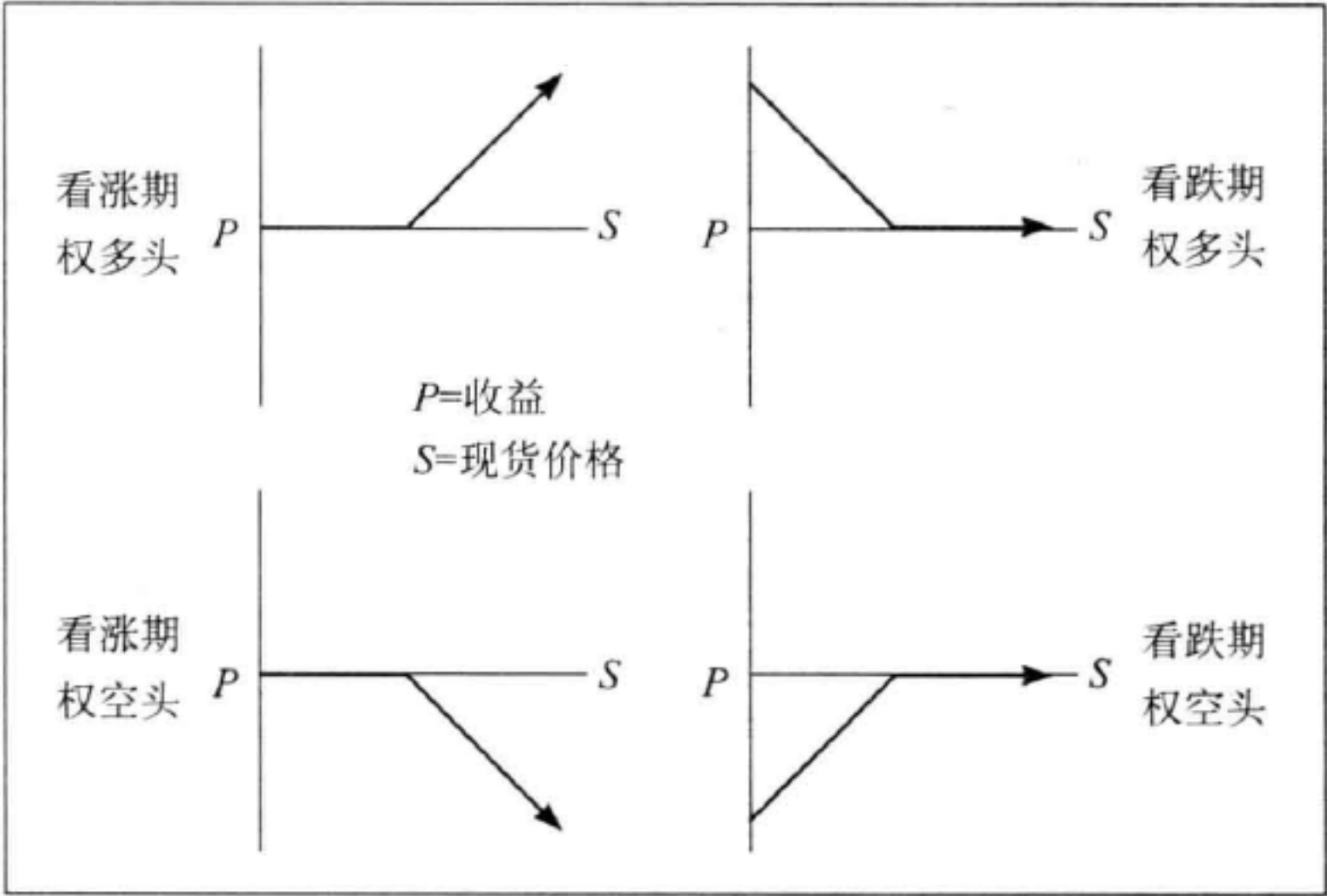


图 3-12 期权收益

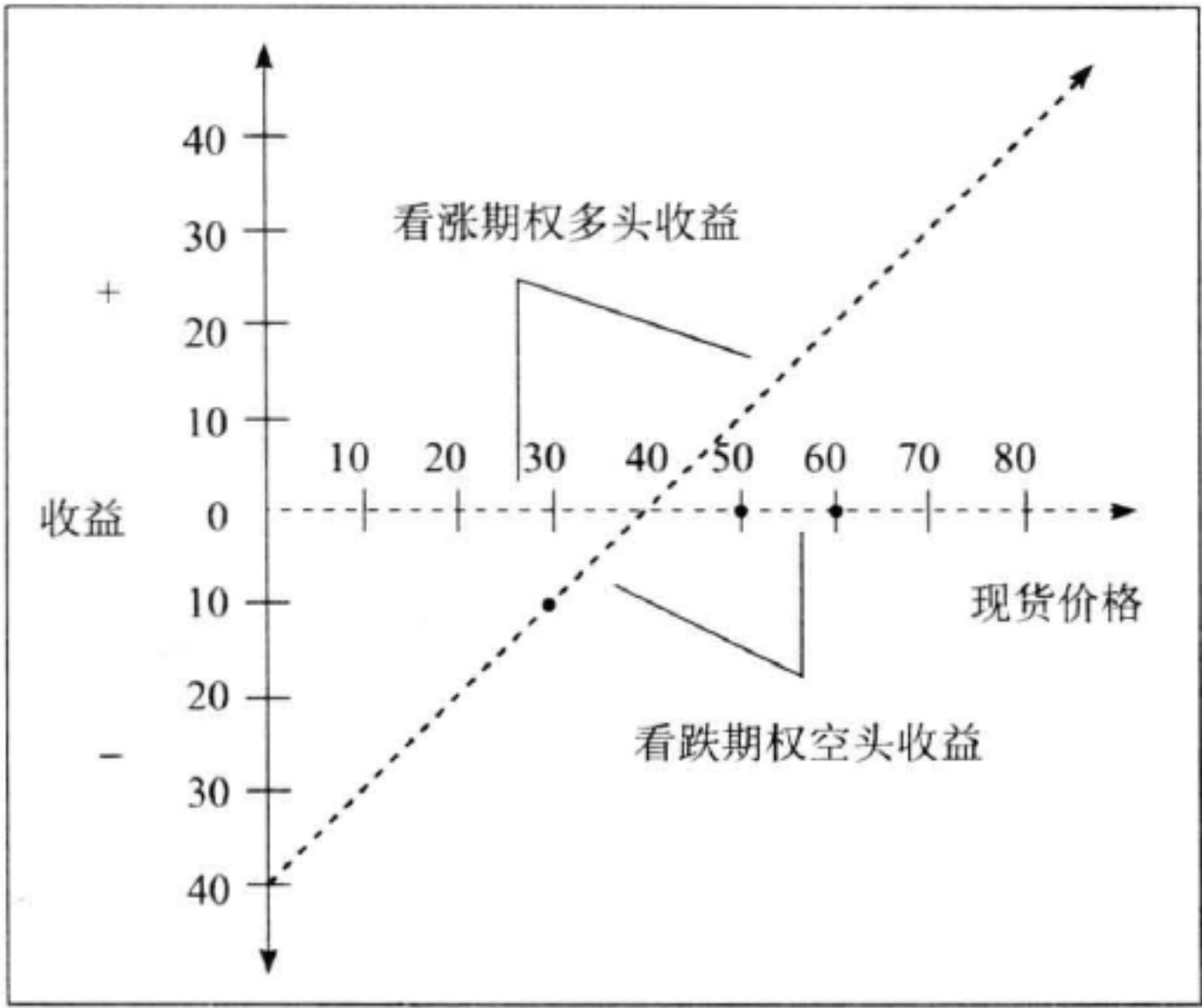


图 3-13 合成期货多头收益

通过组合一张看跌期权的空头和一张看涨期权的多头，我们合成了与期货合约多头一样的收益。现在我们就有两个投资组合，其工具完全不同但是收益完全相同。一价定律告诉我们，这两个投资组合的价格必须相当。因此，套利者就可以同时观测真实的期货市场 and 合成的期货市场，寻找价格差异，并且买入价格较低的组合，卖出价格较高的组合。

这一类的套利机会还有很多。图 3-14 列出了股票、指数和衍生品之间的关系。其中的每一种关系（通过箭头表示）都展示了权益类大超市中的一种潜在套利机会。

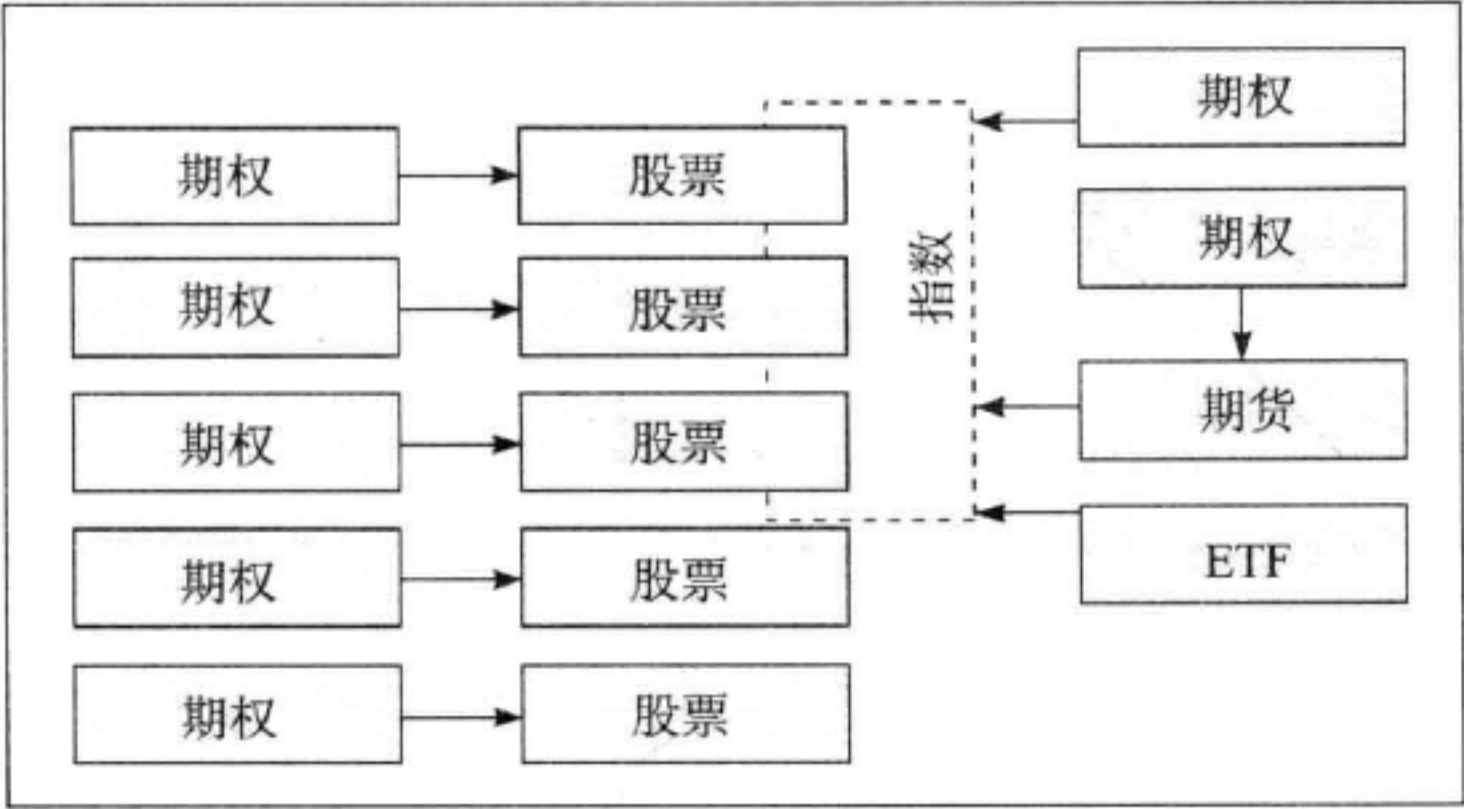


图 3-14 套利机会

期权之间的套利

我们已经接触了一些期权套利策略，还有其他一些常见策略值得我们注意。另外，很多用于股票的套利方法对期权同样适用。例如，不同交易所的交叉市场套利——一个交易所某证券的申卖价低于另一个交易所同一证券的申买价——不论对股票还是期权都起作用。

波动率套利

期权定价公式通常都含有以下几个变量：股票价格、执行价格、

到期时间、利率和波动率。如果标的物是分配股利的股票的话，还必须加入预期股利。这些输入变量中，最有意思的还是波动率。这是因为它是一个不可观测的量。然而，波动率又非常重要，期权交易者在评估期权“价格”的时候，不仅要看期权的美元价格，还会看它的波动率水平。虽然不能直接观测到波动率，但你可以在给定期权价格和其他变量的条件下，通过反解期权定价公式把波动率给算出来，此波动率即为**隐含波动率**（implied volatility，IV）。因为你首先得知道期权价格才能推出波动率，咋看起来这种技术对计算期权价格没有什么太大的帮助（想象一下，有一个教怎么做鸡汤的菜谱把鸡汤作为其原料之一）。但是实际上这种方法相当有用。如果计算出一系列期权的IV（不同执行价格、不同到期时间），你就可以把这些波动率水平画在一个三维的网格上，这就是所谓的波动率曲面。图3-15显示了波动率曲面看起来会是什么样子。

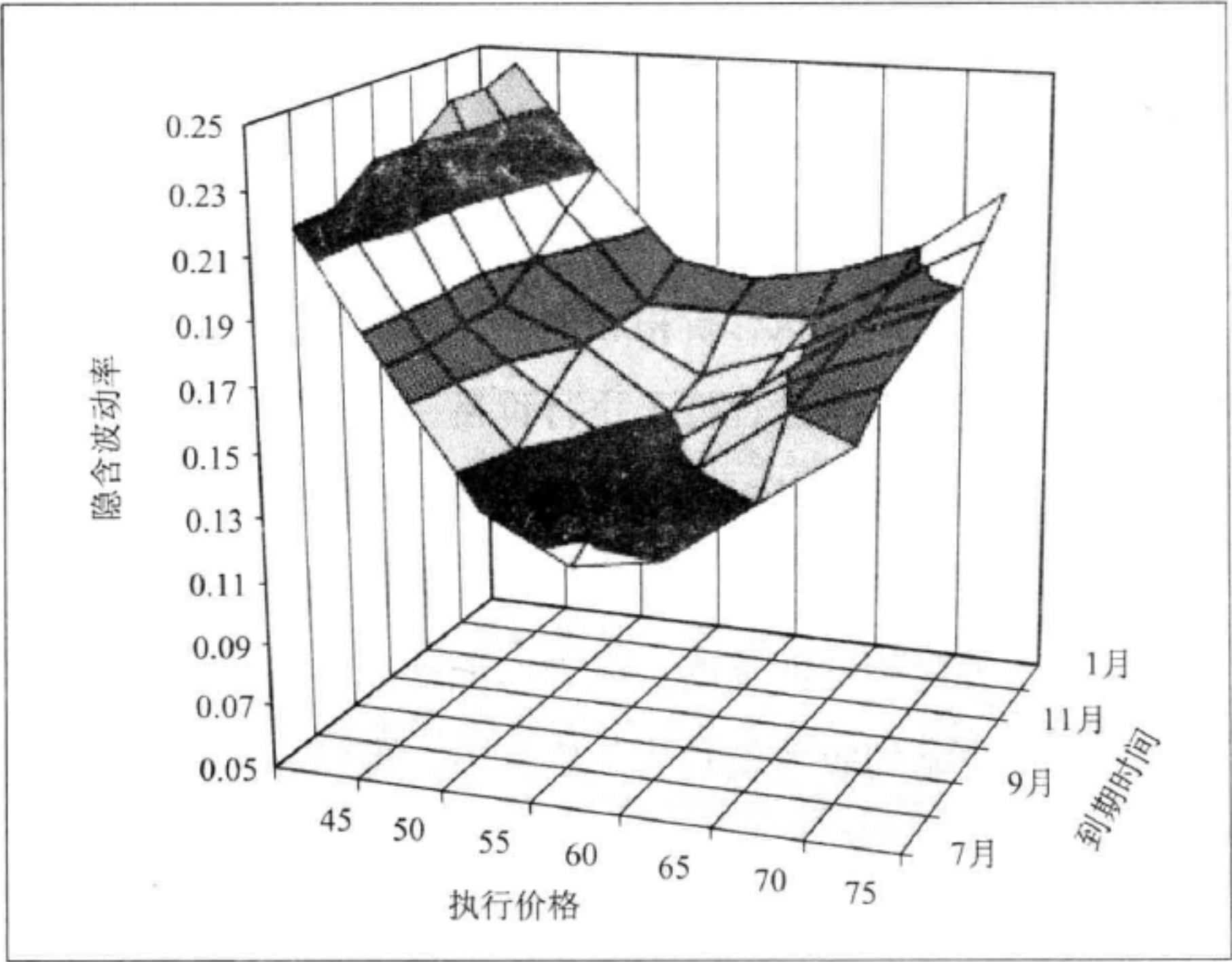


图 3-15 波动率曲面

这个曲面有一个轴是到期时间，另一个轴是执行价格，第三个轴是隐含波动率。曲面可以相当可靠地告诉你任何一个给定的期权合约的波动率应该是多少。如果你看到一个合约的波动率不在曲面之中或者曲面附近，那么这个合约要么是定价过高（如果波动率比曲面高），要么是定价过低（如果比曲面低）。如果定价过高，你可以卖出这份合约，并且在合理价格买入一份合成期权。如果定价过低，操作正好相反。这里说的东西有所简化，但是其要点就在于波动率是一种找出错误定价期权的手段。

价差套利

前面我们看到了四种单个期权的基本收益曲线，即看涨期权多头、看涨期权空头、看跌期权多头和看跌期权空头。大部分期权都是单独交易的，它们的收益是这四种曲线中的一种。但期权也可以两个或者更多个在一块组合交易，以形成一些其他类型的收益曲线。这种类型的交易称为价差（spread）或者策略（strategy）。图 3-16 显示了一种非常简单的策略，即跨式（straddle）策略。

这个策略由一个看涨期权的多头和一个看跌期权的空头组成，这两个期权的到期时间和执行价格相同。你可以看到，不论是到期时股票价格远远高于还是远远低于执行价格，跨式策略都可以产生正收益。这个策略的持有者对股票价格上涨还是下跌是无所谓的，他只是希望价格能够随便在哪个方向有大幅运动。因为这个原因，跨式策略是一个方向中性的策略，或者叫波动率策略，因为标的股票的波动越高，策略的收益就越大（后文中，我们会看到一个用跨式策略降低波动率风险的例子）。

价差组合的价格等于该策略每个期权（或每腿）价格（或权利金）的简单加和。在图 3-16 中，收益曲线垂直移动的距离表示权利

金。从中可以看出，看涨期权这一腿的权利金大约是 5 美元，看跌期权这一腿大约是 6 美元，因此总共权利金在 11 美元左右。换句话说，如果期权到期时，标的股票的价格正好等于 K （执行价格），那么这个跨式的买家就会损失 11 美元。但只要股票价格高于或者低于 11 美元，损失就会小一些。并且，如果股票价格到时离 K 非常远——不管在哪个方向——跨式策略就赚钱了（如果是跨式策略空头，那么收益就要倒过来并且垂直向上移动）。

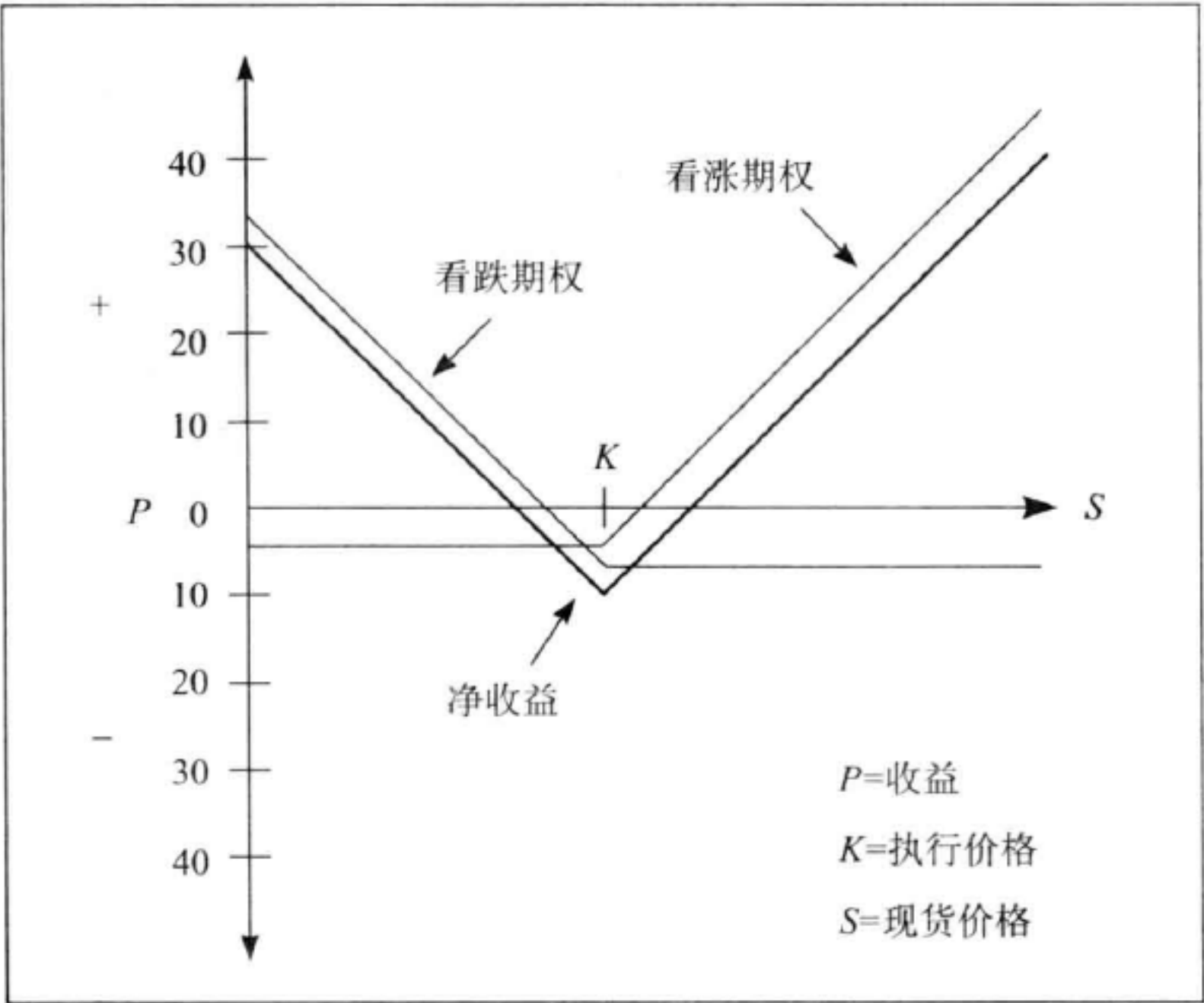


图 3-16 跨式多头

很多价差组合（像跨式、看涨价差、蝶式等）的使用频率非常高，它们也有自己的报价和价格，就像这些组合都是一个单独的证券似的，虽然实际上它们是多个证券的组合。因此，比较大的期权交易所会给价差提供一个单独的委托单簿。例如，在 CBOE，这个委托单簿称为复杂委托单簿（complex order book，COB）。如果想买图 3-16 中的跨式价差，你或许可以在市场价差委托单簿中看到

11.40 × 11.50 的报价。要是觉得这个价格还不错，你可以打掉申卖，支付 11.50 美元，从而得到一个包含一腿看涨和一腿看跌的组合头寸，见图 3-16。

显然，价差组合的价格在数学上由其每腿期权的价格所决定，前者必须等于后者之和。如果不是这样，一价定律就告诉我们出现了套利机会。回到图 3-16 的例子，我们已经知道可以在价差委托单簿中以 11.50 美元的价格买到跨式组合。如果看涨期权委托单簿的市场报价是 5.20 × 5.25，看跌期权是 6.20 × 6.25，那么你还可以分别提交两张委托单，一张买入看涨，一张买入看跌[⊖]，这样你就以 11.50 美元（5.25+6.25）的价格（在所谓的自然市场中）买入一份一模一样的价差组合，其价格与价差委托单簿中的价格也是一样的。这种情况是一价定律所说的理想情况。现在设想一下，看涨期权的市场报价升到了 5.35 × 5.40，但是看跌期权市场仍然是 6.20 × 6.25，并且跨式价差市场也保持在 11.40 × 11.50。你看出其中的套利机会了吗？如果足够快的话，你可以在跨式委托单簿中以 11.50 美元的价格买入跨式，并在自然市场上以 11.55 美元（5.35+6.20）的价格卖出去，从中赚到 5 分钱。自动价差套利还是一个相对较新的领域，因为交易所刚刚提供做这些事情所必需的价差委托单簿电子接口。随着越来越多交易所的跟进，你可以确定会有更多的电子化套利者对此鹰视狼顾。

预测者策略

我们的套利者寻找一个证券相对于其他一些证券错误定价的情

⊖ 通过自然的方式交易价差有一个风险，就是可能一腿成交了，但另外一腿没有成交，这就是所谓的单腿风险。然而，不少交易所允许你指定一组委托单属于一个价差组合，他们只会在所有各腿委托单都能成交的情况下才成交委托单组合，从而降低了这种风险。

形（例如，违反了一价定律），而预测者（也称为量化交易者或统计套利者）则在时间的尺度上去发现价格的不合理。大体上来说，预测者分析市场数据，去寻找一些场景他可以比较有把握地认为一只以价格 Y 交易的证券会随着时间的推移涨到 Z ，他会以 Y 的价格买入并且等到价格达到 Z 时卖出。或者是证券会从 Y 跌到 X ，他会以 Y 卖出，并以 X 买入平仓。这中间的持仓时间可长可短，范围很大。因为主要是对高频交易感兴趣，我们会关注那些交易最多几秒几分钟价格变动的策略，而不是几星期几个月的策略。

配对交易（第二部分）

毫无疑问，我们早先提到过的配对交易是时域上错误定价的最为基础的例子，并且其时间跨度非常之短。在这个例子中，有两只股票的价格走势通常前后呼应，如果一只价格已经变动了，另外一只没有跟上，那么这只股票就被认为是滞后的。预测者认为这只股票最终会跟上前面那只股票。但在它还没跟上之前，预测者就认为它被错误定价了。在统计套利的术语里，这段时间被认为是有 α 的。 α 是用来衡量超额收益的，预测者常常用它来表示错误定价的大小。一句笼统的“那只股票价格被低估了”，即使这句话说得再对，在这里也用途不大。我们需要精确地知道这只股票到底低估了几分钱，或者是零点几分钱，因为我们已经把这些实际工作交给电脑了，而电脑需要非常明确的指令。 α 就是起到这个作用。

一旦发现了所谓的 α ，并且算出了具体数值，通常就可以用一个名词叫做微观价格（microprice）来表示它。这是一个预测者认定的公平价格，在这个价格之上，无论买卖都无利可图。例如，在某个时刻，微软公司股票的微观价格可能是 30.4971 美元。过一会儿我们发现它的价格相对滞后了（比如，相对于谷歌公司的股票），它

的 α 可能会增长，其微观价格也就上升到比如说是 30.5112 美元。

微观价格怎么用呢？记住，所有的高频交易机会都来自于有且仅有一个地方，即委托单簿。预测者的眼睛——我们称之为他的电子眼——时刻紧盯着申买和申卖的每次变化。他根据自己的 α 策略，寻找过高的申买价和过低的申卖价。当然，他需要一个价格基准来与这些申买和申卖价进行比较，以决定吃掉哪张买单或者打掉哪张卖单。这个基准价格就是微观价格，它把预测者在委托单簿中的工作简化成两个基本规则：如果申买价高于微观价格，吃掉它；如果申卖价低于微观价格，也吃掉它。

这里面还有其他一些小细节，比如买价和卖价需要超过微观价格多少才交易，并且预测者也不是只有在买价和卖价变化时才进行比较，微观价格变化时也在进行比较。但是，和金融领域的其他很多东西比起来，这些都还算是小菜一碟。然而，这也正是微观价格方法的美妙之处，它让电子眼的工作——也就是它的算法——非常的简单。对算法而言，简单就是好，这是因为它让程序员可以写出十分高效和快速执行的代码。再把算法想象成一张菜谱，下面哪一个你得花更长的时间准备呢：一份煎蛋三明治，还是查理特罗特餐厅的冷冻金樽鱼加上带有柑橘腌三文鱼和酢浆草的加州小龙虾？

期货延迟

这个策略与期货对股票篮子套利策略有类似之处。然而，在套利策略中，我们交易相对于指数期货合约的一篮子股票。在这里，预测者将指数的成分股分开考虑，计算各只股票由于滞后指数期货而产生的 α ，并且分别进行交易。这个策略与简单配对交易策略也有类似之处，只不过这里的股票并不是与另外一只股票配对，而是与其所在指数的期货合约配对。

让我们回头考虑一下芝加哥商品交易所交易的标普 500 指数迷你期货合约 (ES)。这个合约的流动性非常好，它每天的成交金额大约有 1 500 亿美元。注意，是每一天！这基本上是整个股票市场成交金额的 2 倍。因此，ES 被许多人认为是美国股市名副其实的风向标。ES 期货市场和其他期货市场一样，具有为其标的基础产品进行价格发现的功能，这里的基础产品就是美国最大的 500 只股票。换句话说，当投资界认为现在是买入美国股票的好时机的时候——不是哪一只股票，而是整个美国股票市场——他们往往会从交易股指期货开始。例如，当对市场有重大影响的消息突然出现时，那么最先对之做出反应的往往是 CME 的 ES 合约。

不过这看起来似乎有些倒因为果了，因为标普 500 指数应该是反映 500 只股票平均价格的，而不是这些股票要反过来适应指数。但由于某些原因（这些原因不在本书的讨论范围），预测者知道，整个市场的价格变化一般首先体现在 ES 合约上，然后才是各个成分股。另外，由于有大量的历史数据，预测者可以相当准确地算出当 ES 的价格变化时，这 500 只股票每只的价格会变动多少。不同股票对整体市场变动的敏感程度通常是不一样的。^①

例如，他可能知道，在某种情况下 ES 的中间价上涨 1%（例如，上涨发生在 1 秒钟之内），微软公司的股票（MSFT）会上涨 0.8% 或者是 80 个基点。如果下降 1%，MSFT 会下降 80 个基点。依此类推，他可以算出 500 只股票每一只的情况。现在，我们要做的就是观察 ES 的价格变化，计算出 500 只股票每一只的微观价格，并用这些微观价格和这只股票每个市场的委托单簿进行比较，以发现交易机会。如果发现有申买价比微观价格高，则吃掉该买单。如果发现有申卖

① 在资本资产定价模型（CAPM）中，这种敏感程度称为股票的 beta。

价比微观价格低，则打掉该卖单。等股票价格调整到位，就可以平掉仓位并获得相应的利润。

那么在 ES 的价格变动之后，成分股价格变动之前，预测者有多长时间来计算这 500 只股票的微观价格呢？如果他不能在毫秒——甚至微秒——量级之内完成这件事的话，那他就没有什么必要费神尝试了。这个策略广为人知，很多公司已经进行了大量投资来抢占先机。

这个策略不仅可以用在滞后的股票上，还可以用在其他指数期货上。例如，经过对市场数据的仔细分析，预测者可能会发现 ES 和纳斯达克 100 迷你期货合约（NQ）之间的领先-滞后关系，或者是和道琼斯迷你期货合约之间的关系，他就可以和 500 只成分股一样，训练他的电子眼来发现这些市场上的 alpha。

事件驱动

这里的基本想法就是，对那些很有可能会对股票价格有影响的事件，要第一个发现并对之做出响应。比如，有一家公司已经很多年的季度分红都是 20 美分，突然宣布将本季度的分红削减为 15 美分。或者是突然宣布它们未能实现季度预期收入，或者是超过了预期收入。这些消息会让每个人都大吃一惊。在这里，未预期的因素是十分关键的。如果每个人都预期到了这家公司要削减分红，或者是不能达到预期收入目标，那么在事件实际公布之前，这些预期就已经反映在股票价格里面了。

如果不考虑内幕交易的话，一则新的重大的突发消息是能够影响到价格运动的。例如，分红降低的消息会降低你持有股票的预期收益，因此会降低股票价值。如果你能在消息公布时，抢在所有人之前卖空股票，你可以预期接下来股价会下跌。等到股票真的跌下来，

你可以买回股票平仓获利。如果股利有一个未预期的上升，你可以买入股票等到价格上涨之后再卖出。

期权做市商也得对股利发放的消息特别留神，这不仅是因为他们要利用这些消息，还得防范反被其他人利用。计算期权价格时自然而然必须考虑股票的当前价格。例如，一份执行价格为 10 美元的看涨期权，在股票价格为 12 美元时，其价格起码是股票价格为 11 美元时的 2 倍。这一点是很明显的。现在，我们可以把股利看成是股票“漏”出来的钱。[⊖] 因此，如果一只股票在其期权到期之前支付股利，其期权的价值要比这只股票不支付股利时低，其价值差异取决于股利水平。如果股利有变化，期权的价值也会有变化。因此，如果你能对股利变化第一个做出反应，你就可以像在股票市场上一样，抢在期权价格变化之前进行交易。并且，如果你在为期权报价的话，你会想尽可能快地修改你的报价，以免被那些事件驱动的交易者给捡便宜了。

当然，只有在很少数的情况下，所公告的股利或者收益是大家都没有预期到的。即使没有内幕消息，还是有些紧密跟踪公司基本面的分析师能够通过公司管理状况等找到一些信号，对公司的公告做出合理预测。但即使是分析师，也时不时会大吃一惊，当他们都感到惊讶的时候，事件驱动交易者就会抢在所有人前面，对未预期的价格变动进行交易。

趋势跟踪

如果你能精确地预测未来的（比如下周）股票价格，那么只要你愿意，你就能想赚多少钱就赚多少钱。股票价格不是要涨了吗？那

⊖ 我在芝加哥大学的教授黄明常常这么说。

就买入然后等着卖出。要下跌了？那就卖出然后在低位买回来。当然，没有人能做到这一点，我们的预测者也无此奢望。然而，他还是相信，在极短极短的时间段里面，是有可能发现价格趋势的。并且他也知道并不需要每次都对。如果获胜的概率能比一半高一点点，并且在预测错的时候能够有办法活下来，那么赚钱也就是迟早的事了。

那预测者究竟是怎样发现趋势的呢？能够预测多远呢？不好意思，这个领域的成功者都没啥兴趣写一本书来谈一谈，或者是当一个作者问他的时候把秘诀分享一下（这一点你得相信我）。但是据说在股指期货市场上，至少有一种曾经有效的方法。其想法就是观察不同时间周期的移动平均线，比较短周期移动平均线的值和长周期移动平均线的值。例如，每当股票价格有变化的时候，你都重新计算一下两个平均值。一个是过去 5 分钟所有价格的平均值，另一个是过去 5 个小时所有价格的平均值。如果 5 分钟的价格平均值低于 5 小时价格平均值，就说市场有一个向下的趋势。如果情况相反，就认为市场趋势是向上的。这个方法完美吗？当然不是。但还是那个说法，你只需要对的次数比错的次数多就可以了。

另外还有一种趋势跟踪的策略称为**动量策略**——这可能是一个从流体力学或者是物理学其他分支中借用过来的概念。这种方法用一个模型来拟合价格变动，以便在趋势发生之前，以一定的可信度来预测趋势。就像大自然的很多现象（天气变化、水龙头的水流等）一样，也不难通过查看一张价格图表（比如，标普 500 的价格图）来得出一些基本并且合理的（如果不是绝对正确的话）结论。例如，股票一般不会瞬间从一个价格跳到很远的另外一个价格。股票当然可能从 5 美元涨到 30 美元，但这不会是一笔发生的。这中间一般都

有一个路径，即使可能很短，但也是一个路径，就像天气不会一瞬间就从瓢泼大雨变成艳阳高照一样。

有时你可能会发现一波价格上升之后倾向于有另一波价格上升，一波价格下降之后倾向于有另一波下降。当然，现实并不会总是这样，但是可能你的分析告诉你这种情况发生还是相当多的，或许就是在次数上占据少许的优势。多年以来，科学家一直在将自然现象拟合到合理的预测模型之中，并且不停地修正和改进这些模型，虽然这些模型永远也不可能达到完美。预测者在这里也可以使用相同的思想。当然，这并不是一件容易的事情，相反，即使只是找到一个比纯粹赌博好一点的策略都难得要死，尤其是现在有那么多公司都寻找相同的东西，这就难上加难了。但这也不能阻止一些公司去尝试，并且偶尔也会有人成功。

均值回归

我们已经看到由于供需不平衡所造成的价格动力现象是如何影响投资者和做市商的，实际上，这些现象对于预测者来说也十分有用。申买委托单的增多会推动价格上涨，上涨幅度与委托单的相对增量有关。申卖委托单的增加具有相同效果，只不过方向相反，其作用是推动价格下降，下降幅度也与委托单的增量有某种关系。当然，这种不平衡状态不会永远持续下去，它会逐渐消失。然而，它好像总是要先走过头，然后再反弹回到应在的位置。也就是说，市场常常对买卖不平衡现象反应过度，然后再进行自我修正。即价格会回归到平均值，或者，大致来说，回归到市场买卖中间价。当然，价格不一定会完全回归，尤其是市场的某一边有持续压力的时候更是如此。但是这种均值回归现象在统计上是足够显著的，因此，我们的预测者对此非常感兴趣。

再假设在 $500|1.00 \times 1.10|500$ 的市场上，在申买这边出现了一只 5 000 股买入的大象。有可能在大象完全成交之前，市场价格会快速上涨到 1.15×1.20 ，但是预测者知道，这个价格只是暂时的。在市场达到最高峰的时候，它是有 α 的。比如说，预测者算出的微观价格是 1.10 美元，那么他就会疯狂地吃掉 1.15 美元上的申买单，在这个价位倾其所有进行卖出。当市场回归到均值的时候，比如，回归到初始的 1.00×1.10 ，他就会高高兴兴地打掉 1.10 美元上的申卖，买入平仓，并且干净漂亮地赚到 5 分钱。因为大象可能还会继续买入，所以他会迫不及待地等到大象回来之前完成空头平仓操作。需要提醒一下，这是一个完全虚构出来的极端的例子，但是它说明了均值回归的基本思想。

高频交易者的起源

之前，我们把高频交易者描述成一个特殊的做市商，他利用超快的执行速度和极强的预测能力来赚取利润。例如，你常常可以看到高频交易者以微观价格为基准进行买卖报价，并且根据手中的存货调整买卖价格的倾斜度。 α 可能来自于趋势跟踪、均值回归、配对交易或其他一些预测策略。趋势跟踪策略尤为引人注目。考虑一个 1.04×1.05 的市场，高频交易者比较有把握接下来会有一个下行趋势，比如说到 1.00×1.01 。他当然可以吃掉 1.04 美元上的买单，等到市场跌到 1.02×1.03 ，再打掉申卖平掉空头仓位，并获得一分钱的利润。因为知道趋势还没有结束，他还可以把上面的事情重复一遍，他可以吃掉 1.02 美元上的申买，等申卖跌到更低的价位。起码他可以等到申卖跌到 1.02 美元，这样他仍然可以赚到一定的返还。

直到2009年的前几个月，人们都在舔舐伤口的时候，高频交易者才像超新星爆发一样突然出现的公众的视野中，但实际上，高频交易的发展并不是一蹴而就的。他们由传统的做市商经过多年演化而来，起码有两个趋势在这个过程之中起到了非常重要的作用。

第一个趋势与越来越多的买方公司采用计算机处理大额委托单有关。曾经，机构投资者（比如共同基金经理的经纪商）可以给他认识的几个关系不错的做市商打一轮电话，私下里告诉他们要交易一笔大额委托单，并且他相信这些做市商交易员不会抢在这个大额委托单前面进行交易。为什么他这么有信心呢？因为这些交易员需要维持自己的声誉。如果一个卖方交易员接到经纪商的电话，知道有一笔大额委托，他说“不了，谢谢”，接着就挂了电话并在市场中交易使得价格朝不利于经纪商的方向移动，那么可以想见，这个交易员以后就不会再接到这种电话了。他知道，这样做会威胁到自己的生存。但当越来越多的经纪商开始使用计算机处理委托单的时候，声誉风险就消失殆尽。例如，如果一个做市商可以探测到冰山单的信号，并且把它朝对自己有利的方向牵引，这事又有谁知道呢？在其他公司也学会了探测隐藏流动性的模式识别方法之后，游戏就从怎么探测到流动性变成怎么比其他人更快。自动化的流动性探测成了一个公平游戏，计算机装备竞赛也由此展开。

第二个有助于解释为什么高频交易者从传统做市商中发展而来的趋势与一种称为*逆向选择*的现象有关。这里的基本逻辑是，当市场参与者拥有的信息或能力彼此不对称时（有些人知道的比其他人多一些，或者技术比其他人好一些），市场会对有更多信息或能力的人更有利。证券市场中的“知情交易者”问题就是一个例子。试想一下，做市商的对手是一个知情交易者。他可能是通过仔细而不厌其

烦地分析公司收入、管理计划或者类似的一些公司状况，得到了做市商所不知道的信息。或者他可能是一个量化交易者——即我们的预测者——通过高速运转的计算机进行最为先进的统计分析。他可以基于很强的 α 信号做出准确的预测，因此他可以在市场中看见别人不知道的信息，或者至少是比其他人先看到这些信息。传统的做市商无知者无畏，可能给出了 1.00×1.02 的报价，而具有信息优势的交易者知道股票的价值实际上应该是 1.03 美元。

这种情况你懂的。那么做市商会长期容忍下去吗？当然不会。当买卖价差窄到一分钱或者更少的时候，知情交易者在信息上的小优势就成了缺乏信息的做市商的大麻烦。最起码在交易最活跃的股票上，做市商只有像预测者一样精明才能赚到钱。既然你不能打败他们，那就加入他们吧。

第 4 章

All About High-Frequency Trading

提 速

即使是最好的高频交易策略，在实现之前也无法为你带来收益，所以你还需要在策略的实现上多下些工夫。事实上，一个构建得不好的高频交易系统可能比没有高频交易系统还要糟。对你而言更糟糕的是，那些已经拥有了经历过优化和实战考验的交易系统的公司已经在市场上等你了。潜伏在水中游弋的鲨鱼们很乐意看到小鱼们出现在自己的水域，对它们来说，这意味着可以美餐一顿了。

在本章中，我们将讨论一些新的策略，它们并不是交易策略本身，而是通过设计并实现一些定制的计算机软硬件解决方案（即构建高频交易系统），将交易策略从理论变成现实。简而言之，高频交易系统就是一系列的计算机服务器和运行在其上的软件，它们监听市场数据，在其中搜索能带来利润的交易机会，并试图抢在别人之前将这些机会转化成利润。在之后的第 5 章中，我们将描述一个假想的高频交易系统的具体框架，并揭开高频交易系统神秘的面纱，如果你愿意的话。

需要指出的是，我们在这里着重指出技术问题的重要性，并不代

表之前花了大量笔墨讨论过的交易策略就不那么重要了。毕竟，交易策略能赚钱本身就是它存在的价值。同样，如果把高频交易比作一台挤奶机，强调技术这条腿的重要性，也完全不意味着对另一条腿——数学地位的削弱。策略、技术和数学这三条腿之间互相隔着120度，不管哪条腿短了，高频交易这台挤奶机都站不稳。那些努力地高频交易系统的数学基础工作的人称做量化分析师。如果在这个游戏中，你玩不转数学，那么你已经出局了。事实上，你得对付的数学问题还不少。量化分析师既需要处理一些简单的问题，比如计算一串未来的预期股利的现值、在拆股或者合股之后重新计算股票期权的价值之类的；他们还需要处理一些更困难的问题，比如最大化多维相关系数矩阵的计算效率、在许多不同形式的有限差分方法中选择最好的一种，等等。虽然他们在工作时用来喝咖啡的马克杯上不太会印上布莱克－斯科尔斯偏微分方程之类的东西，但他们早就对此烂熟于心了。[⊖]

事实上，我们讨论的是一个关于人的天赋和能力的话题，虽然在高频交易系统的构建中，我们不能绝对地说计划和实施团队的能力是决定高频交易系统成败的最关键的因素，但如果说构建成功的高频交易系统的秘密——就像无数其他复杂的任务一样——被印在

⊖ 其实你也能做到这一点。布莱克－斯科尔斯偏微分方程的一种形式是这样的：

$$\frac{\partial f}{\partial t} + rS \frac{\partial f}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} = rf$$

根据布莱克、斯科尔斯和默顿的结论，在一些给定的假设条件下（例如：标的股票的波动率为常数等，虽然在现实中这个条件显然是不成立的，但我们在这本小书里不对它作更深入的讨论），该方程对于一只标的为股票的衍生品是有效的。需要顺带说明一下的是，布莱克－斯科尔斯偏微分方程并不等价于著名的BS期权定价公式。BS期权定价公式是通过建立期权价格的边界条件，从一个价格满足布莱克－斯科尔斯偏微分方程的期权之上推导出来的。上式中， f 表示衍生品价格， t 表示时间， S 表示标的股票价格， σ 表示波动率， r 表示无风险利率。方程左边的那些项是偏导数，例如，第一项表示的是价格 f 对时间 t 的变化率，其他的几项也都是是一些类似的数学表达，我们就不作更多的说明了。

一张保险杠贴纸上，那么它上面显然应该写着“蠢货，这全靠人”。构建一个成功的高频交易系统是一个复杂而困难的大工程，它需要参与其中的团队成员有足够的天赋、品格和动力来完成这样的任务，目前那些最好的高频交易系统在构建的过程中都近乎巧合地把拥有这些特质的人集中在了一起。在一个构建高频交易系统的团队里，核心成员们需要想象力、紧迫感、前瞻性，以及恰好足够的自负——但这种自负不能过渡到无法与其他的那些拥有想象力、紧迫感、前瞻性和适度自负的人共处的地步。仅仅雇来一些明星交易员、明星程序员和明星量化分析师并不意味着成功。毫无疑问，这些明星们是很有价值的，但是你还需要能力出众的管理者把他们整合在一起。举几个有限的例子：交易软件需要测试，而对于高频交易系统来说，由于苛刻的时间要求和系统的复杂性，软件的测试和质量检验可能会成为一项让人精疲力尽的工作，并不是每个人都能很好地完成这样的任务。近年来，网络工程已经变得越来越像一门艺术了，也难怪现在那些拥有思科认证架构师这类证书的工程师在这个领域里几乎可以随便给自己开工资了。同样需要注意的是，本书中的内容大部分还是有关所谓的“交易前台操作”的，它涵盖了交易发生时刻之前的一切内容。当交易完成后，后台的诸如清算、合规和交付保证金之类的关键过程才开始运转。这些都是极其重要的工作，它们的价值是不言而喻的。

也许你已经注意到了一种可能性：想要有效地在高频交易领域与对手竞争，一家公司必须构建属于自己的定制的高频交易系统。但从可行性上而言，既然已经有现成的高频交易软件解决方案，只需要购买软件许可就可以使用，诸如 ORC、Actant 和 RTS 这类著名的公司都提供全自动的做市软件，它们几乎能实现你所需要的所有

功能——报价引擎、电子眼、自动对冲等，那么为什么还要自建系统呢？毫无疑问，有些公司的确使用这些产品并取得了成功（否则这些软件供应商早就无法生存下去了）。但那些处在高频交易领域前沿的公司都宁可构建自己的系统，而不依赖于第三方提供的软件。自建系统显然比使用第三方软件贵得多，甚至贵上几个数量级。但是相比那些不愿意进行这项投资的公司，它们也能够因为自建系统而享受到明显的优势，这主要有两个原因，我们接下来将讨论这个问题。

购买还是自建

当你在竞争中和对手使用同样的工具，你的优势至多也就是能比对手更好地使用那些工具。举例来说，两个司机如果开着相同的法拉利赛车在同一条赛道上比赛，他们之间的水平高低也很容易判断出来。但在竞争中，这样的优势也许是微不足道的。但是如果你建造了自己的工具，你就可以从设计的质量和工艺中得到额外的竞争优势。就像我们接下来将会提到的，编写软件的程序员水平对一个高频交易系统的软件算法效率有重要的影响，而程序员能力之间可能存在着很大的差异。高频交易和赛车不一样的是，在高频交易的赛道上，赛车里是没有司机的，交易系统的设计和构建过程就是你证明自己比竞争对手更好的赛场。

高频交易公司选择自建系统的另一个原因看起来很自然——这让他们在任何时候都可以去优化其中的软件。第三方的软件供应商总是会尽量满足你的软件优化和漏洞修改需求，但是他们的客户很多，所以需要把这些需求排排队。那些最富野心的高频交易公司没有任何等待的耐心，他们不会等着软件供应商发布下一次的产品更新，而是希望根据自己的时间表来修改漏洞并提升系统，要做到这一点，

就只能自建系统。任何一个在交易系统开发领域工作过的人（即使是在高频交易出现之前）都知道在产品更新之间，你会常常需要用到“即时补丁”和“紧急漏洞修复”之类的东西。

当然，在某些时候，这种“我们必须建造自己的系统”的想法可能是有点过头了。一些聪明的高频交易公司会选择只是自建一部分的系统组件，而非整个系统。举例来说，自建一些所谓的“交易基础设施”所能带来的优势就是非常有限的，数据库就是这种“交易基础设施”的一个典型例子。同样，我们接下来将要讨论的消息组件也是一种典型的“交易基础设施”。

组合产品

我们已经讨论过套利者是如何利用不同的证券和投资组合的价格关系的。理所当然，高频交易系统的设计者在最大化策略实现的效率时，也需要将这些价格关系考虑在内。在期权投资中，这些价格关系可能带来的套利机会尤其丰富。我们考虑下面这些合约：

- SPX——这是一种在芝加哥期权交易所（CBOE）交易的欧式期权[⊖]，其标的是当前的标普 500 股票指数。SPX 是一种现金期权，如果在到期日期权处于实值[⊖]状态，会根据当时的标普 500 股票指数水平和执行价格之差进行结算。

⊖ 一个欧式期权只能在事先约定的到期日执行，而一个美式期权则能够在到期日以及之前的任何一天执行。如果其他条件相同，美式期权的价格要高于欧式期权。需要注意的是，它们之间的区别与地理并没有什么关系。尽管如此，当人们看到“百慕大期权”这个名字的时候，还是会不可避免地把它和那个地理位置联系起来，这是一种可以在到期日以及之前的一系列约定日期执行的期权，其他条件相同时，它的价格应该在欧式期权价格和美式期权价格之间的某个位置上，就像百慕大也处于欧洲和美洲之间一样。

⊖ 在期权存续期内，“价格状态”描述了标的物现价与执行价格之间的差异。对于一个看涨期权而言，如果执行价格低于标的物现价则处于实值状态，如果执行价格高于标的物现价则处于虚值状态，如果两者相等，则处于平值状态。

- XSP——称做“迷你 SPX”，是一种同样仅在芝加哥期权交易所交易的欧式期权，与 SPX 的不同之处在于，它的标的是标普 500 股票指数的 1/10。
- ESO——这是一种在芝加哥商品交易所（CME）交易的美式期权，它的标的物是一份 E-mini 标普 500（简称 ES）期货合约。如果 ESO 期权合约在到期时处于实值状态（例如，看涨期权的执行价格低于当前的 ES 指数），期权买方会获得相应的 ES 期货头寸。
- SPY——这是一类目前在美国所有的交易所都能够交易的美式期权合约，它的标的物是 Spider ETF 或者跟踪该指数的股票。当期权买方在实值状态下行权时，获得的是相应的 ETF 份额。

这四种合约的相同之处在于，它们的价值都是从标普 500 股票指数衍生出来的。给定其他条件，这四种合约的理论价值都会随着标普 500 指数的变动而变动。除了这四种合约以外，市场上还有很多种与标普指数相关的期权合约。这些合约中，由于结构的不同，它们的价值随标普指数的变动方式也不尽相同。例如，XSP 的标的是标普 500 指数的 1/10，而 SPX 的标的则是该指数本身。此外，期权的类型（美式或欧式）也必须考虑在内。不过处理这些数学上的小细节，对“宽客”们（quants）来说是相对轻松的任务。

那么该如何利用它们的这种共性呢？一种显而易见的方式是共享波动率曲面。在图 3-15 中描述了一种数据结构，不同执行价格和到期日的期权价格所对应的隐含波动率组合被维护在一张三维图中。除了标的指数的当前点位，当前的指数波动率水平是另一个在期权定价中极其重要的参数。很显然，你只需要维护一张标普 500 指数

的波动率曲面，就可以用它来驱动以上各种期权合约的定价过程。

除了定价之外，这些期权合约都基于标普 500 指数这一共性在风险管理领域也大有用武之地。举例来说，如之前所提到的，一个期权做市商并非总是有机会将交易一个期权的风险通过在同一个合约上的反向交易对冲掉。最难以对冲的风险是 vega，它代表期权价格对标的物波动率变化的敏感程度。就像计算期权的理论价格一样，计算像 vega 这样的期权价格对各种变量的敏感性（希腊字母）同样需要波动率这个参数。通过维护一张共同的波动率曲面，做市商能够在持有这一系列的期权头寸的组合时只关注组合的单一 vega 值，从而通过对一种合约的买卖操作来对冲另一种合约头寸的风险。举个例子，如果你刚刚卖出了一张 SPX 的看涨或者看跌期权，就完全可以通过买一些 SPY 的看涨或看跌期权来对冲掉它的风险。

让设计适应变化

最好的高频交易策略专家都知道，任何策略有效性的半衰期都不是很长，所以用于交易的策略必须不断改进和更新。类似的想法在高频交易系统的设计中也是有效的。所有的软件工程师都明白，有些软件在设计中为将来的拓展预留了空间，而有些在设计时则没有这么做。大多数工程师会在设计中为两个目标而努力——模块化和低耦合。模块化的含义很简单，它是指在建设一个庞大的系统的过程中，首先设计并建造许多相互之间相对独立的小模块（虽然在设计时仍然不可避免地会考虑到它们之间的关联），之后将这些小模块关联在一起构成整个的庞大系统。而耦合性是指系统模块之间的交互，或者说通信的方式。如果这些通信接口简洁且开销很小，那么

这些模块就被称为低耦合的。在这个概念的另一个极端，高耦合的模块之间需要相对来说很复杂的通信机制，所以它们之间相互依赖的程度很高。为了让系统能适应变化，设计的目标应该是一组独立性很高的组件（或者说对象），它们通过一些很简单的接口将各种复杂的内部操作或实现隐藏起来。使用低耦合的组件构造的系统相对来说更容易修改，因为要在这样的系统中添加新功能，你只需要创建并引入新的对象就行了。在系统改造时，这样的架构也能更好地抵御对系统结构的破坏而显得更加优越。

最理想的系统设计目标是——或许它显得有些乌托邦了，在你改造、修复系统中的某一个组件，或是向系统中添加一个全新的组件时，对系统的其他部分不会造成或者只会造成很有限的负面影响。这个目标的反面则是一个非常脆弱的系统，它的内部相互依赖关系太过复杂以至于事实上没人能把它弄清楚，即使一些几乎完全无害的漏洞补丁也有可能使整个系统瘫痪。你决不会想要一套这样的系统，就像你也绝不会想要一辆只是换个前灯就会让刹车失灵的汽车。

模块化和低耦合本身是来源于软件系统设计的概念，但是当你在管理一个构建面向适应未来变化的系统的工程时，这些概念也派得上用场。在这里，方法论是一门艺术，不过大多数人谈起方法论的时候，因为这个词太深，往往都会失去兴趣。人们在使用这些概念时很容易越界，用空谈和大话取代了真实开发中的常识和经验，只要有几年实战经验的软件开发者都会明白我所说的意思。

尽管如此，还是有一些方法和实践能够得到绝大多数人的认可。它们中的绝大多数都与演进式的，或者说递进式的、迭代式的开发方式相关，而与之相对的则是“大爆炸”式的开发方式（这是指一种在庞大的系统开发之前就将与它相关的所有细节规划好的方式）。

在一个演进式的系统开发中，你需要做的第一步是勾勒出一个整体的系统框架。在整个工程开始前，你需要得到整张图画的大致蓝图和终极目标。使用一支可擦写的白板笔来勾画这张蓝图，绝对要比用擦不掉的墨水笔明智得多。一旦完成了一个试验性的设计方案，你就可以开始建造你的系统了，最好是从那些最困难、风险最高的组件开始。你可以先构造出一小部分的组件，将它们放在测试环境中，并诚实而谦虚地评估一下哪些部分能行，哪些部分行不通。然后再回过头去修改它们，当你得到了一个能够运行得比较顺利的系统核心时，就可以向其中添加新的功能和组件了。这种构建系统的哲学是从人们反复从自然母亲身上借鉴而来的：一个真正能够运行良好的复杂系统一般都是在更简单且可行的系统的基础之上演进而来的。

当你努力尝试去设计由低耦合的对象组成的系统架构和能够为伟大软件的演进提供基础的开发流程时，也可能会因为过于执著于这样的目标而走得太远。当然，模块化和渐进式开发这些概念的正确性实在是太显然了，几乎是不证自明的，但是真正的挑战在于，你需要想起那句古老的谚语：“不要让追求完美成为达到优秀的敌人。”回到现实中来，我们还是需要削减支出、支付账单，需要取悦投资者和股东，一个总是停留在白板上的系统设计方案是没有多大意义的。你永远也不可能确定地知道什么时候能够得到一个足够好的架构和开发流程。这些都来自于经验，但即使是经验也无法为你提供确定的答案。当我还在美国银行工作时，虽然我们拥有很多天赋异禀的交易员和程序员，但当我们在构建一个固定收益衍生品的定价系统时，还是会互相提醒，我们的任务只是在基于常规的基础上“把事情搞定”。同样，我们需要在系统开发的过程中不断说出或

者写下“GSD”这几个大写字母[⊖]来提醒彼此：我们的任务是做出好的软件，而不只是构造伟大的想法。

分布式与负载均衡

我们现在已经对由一系列独立的组件构成的高频交易系统有了一个大致印象，在这个系统中，每个组件都有其独特的功能并且能够运作良好：其中的一个组件会监听并重新发布市场数据，另一个组件则会在市场数据中寻找 alpha（也就是能带来盈利的交易信号），还有一个组件负责下单交易，诸如此类。整个系统就是这样被组件化，或者用软件工程的语言来说被分布化了。你会把一项庞大的任务交给 10 个组件还是 20 个组件去完成？这个问题的答案是：模块化的程度不能太低，以致某些模块因为太过庞大而效率低下；同时也不能过高，以致由于模块间的通信压力太大而降低了整个系统通讯机制的效率。接下来会具体地说说这个问题。

我们已经提到的这种任务分配的概念是功能性的（例如，将一项大的任务分解为一系列更小的任务）。另一个与高频交易系统的设计相关的分配问题是，当有一系列相同类型的工作需要完成时，如何在系统的各个组件之间平衡地分配计算量。举个例子：当需要计算你的资产组合中的每一只股票的价格的时候，你很可能会决定开发一个组件，这个组件除了读取每只股票的各种参数，用一些复杂的数学计算把这些参数转换成股票价格并输出之外，别的什么也不干。但是，如果你有 3000 只股票的价格需要计算，你就得好好考虑一下需要一次实例化多少个并行工作的组件对象了。很显然，仅仅

⊖ GSD 是“Get Things Done”这句话的首字母。——译者注

一个组件对象是远远不够的，因为这会导致一长串的任务队列等待处理，你决不会希望看到这种情况出现。当然，你也不需要 3 000 个组件对象，如果这么做，它们在大多数时间都会处于闲置状态，而实例化这么多对象所占用大量的 CPU 和内存资源会大大降低你的系统效率。你需要做几次试验，然后确定一个合适的实例化组件的数目。

除了计算任务的分配，机器资源的分配也提出了另一个任务分配与平衡的难题。每一台服务器所能处理的任务量是有限的（对内存和处理器来说也是如此），如果任务量超过限度，它们就会停滞下来而无法工作。一旦你决定了模块化功能组件的方式以及每个组件对象需要的实例数目，最后的任务就是分配这些服务器上的处理器资源了。同样，通过计划和试验，你可以得到一个合适的分配结果。

这个游戏最后看起来很像是在比赛开一家快餐店。对于一次交易，有一张委托单需要处理，这个过程就像圆面包等着用肉饼和其他的配料填满，薯条等着油炸、包装，饮料等着装进可乐杯，配料等着从仓库里面取出来一样。一个员工显然不足以应付所有这些事情，但是你显然也没必要为每一项任务都分配一个员工专门负责，这些都属于功能性的分配。接下来你还得确定需要的收银机和炸薯条的油锅之类的设备的数目，这些就属于负荷分配的范畴了。最后你还需要确定在快餐店里设几条食品流水线，如果只有一条流水线，你的店里显然会排起长队，人们也不愿意等那么久；但是如果流水线太多，它们的固定成本会太高，大多数时候却是闲置着的，这会让你因为承担了太高的成本而不得不向顾客提价，很多不愿意付更多钱的顾客就会被挤走。

设计最高效的软件

程序员在编写软件来完成某些特定的任务时，通常会在写代码之前通过一些标准去选择软件背后的算法。在高频交易领域，最聪明的程序员都会把效率当做选择算法的最根本标准，并且会不断地花时间去优化算法的效率。算法越高效，处理任务所需要的时间就越短。毕竟我们都知道，在高频交易中，时间就是不折不扣的金钱。

算法，指的是解决某种问题的方法。它并不是一组指令的集合，指令是一种具体有形的东西。而算法则比它要更加抽象，是指令背后的思想。对于任何一个给定的问题，一般来说都会有很多种可能的算法能够解决它。我们举个例子，假设我们想在电话簿里找到某个人的电话号码，我们都知道电话簿里的名字是按姓氏的字母顺序排列的。想得到一个人的号码，我们必须首先找到这个人的名字，而我们有很多不同的办法来做到这一点。

比如我们想要找到 Sara Gruen 的电话号码。一种办法是从电话簿的第一页开始，一页页地翻，直到我们看到“Gruen Sara”这个名字，虽然这种方法可以保证让我们一定找到这个名字，但它实在是太慢了，很容易看出它是一种时间利用率很低的方法。另一种我们很容易想到的办法是根据姓氏的第一个字母，猜测以这个字母开头的名单在电话簿里的大致位置，然后根据翻开页码上名字的开头字母确定该往前还是往后翻。我们大概会从电话簿靠中间的地方翻起，假设我们翻开电话簿后看到的名字是“Mortenson Greg”，我们知道，“Gruen”这个名字按字母排序时在“Mortenson”的前面，所以我们向前翻一些，假设我们又看到一个名字“Eggers Dave”，因为“Gruen”按字母排序在“Eggers”之后，所以我们再向后翻，看到另

一个名字“Lamott Anne”，再往前翻。不断重复这样的做法，直到我们找到“Gruen Sara”这个名字，看看她的电话号码是多少，再打电话给她。^① 计算机科学家把这种方法命名为二分搜索，这种查找的方法很明显比我们前面说的那种顺序搜索效率要高得多。

事实上，高频交易系统的开发者经常需要在一系列的排序和查找算法中进行选择，这些算法的种类很多，算法之间的差异主要在于时间效率、空间效率（占用的内存）以及准确度上。但除了这些排序和查找算法，他们还经常需要设计一些专门针对高频交易应用的高效算法。举个例子，假设一个程序员需要写程序来根据某个代码的标的物^②（比如 SPY）的价格变动来修正期权报价。但是市场上代码为 SPY 的期权合约有接近 2 000 种，其到期日、执行价格和看涨看跌类型都不一样，并且每一种都有独立的委托单簿，作为一个期权做市商，很可能需要维护 2 000 个申买价和 2 000 个申卖价。只要标的 ETF 的价格发生变化^③，即使只有一两美分，做市商就必须以最快的速度更新 4 000 个报价。这必须通过向交易所发送一长串的报价更新请求来完成，每一个更新将精确地替换掉原来的报价。

程序员在处理这个问题时，不仅仅是选出能将这些更新请求以最高的时间效率发送出去的算法就够了，他还需要考虑发送这几千个更新请求的顺序：是从到期时间最短、执行价格最低的合约开始，按这两个维度顺序递增的方式发送请求直到所有合约的报价都得到更新呢？还是从交易最活跃的合约开始呢？又或者从处于最深的实值状态的合约（也就是执行价格最低的看涨期权和执行价格最高的

① 告诉她我们很喜欢她的那本《大象的眼泪》（*Water for Elephants*）。

② 期权“代码”在这里指的是期权的标的股票。

③ 这里需要注意的是，在对 SPY 期权的定价中，Spider ETF 并不是唯一需要监控的标的。一个交易者同样可以通过买卖 ES 期货合约来影响这些期权的价格，因为不论是 ES 还是 SPY，这些衍生品的最终标的物都是标普 500 指数。

看跌期权，它们的 Δ 最大，这意味着期权价格对标的物价格变动最敏感，所以在标的物价格变动时，这些期权的价格风险最大）开始呢？还是应该选择一些另外的算法呢？在这个问题上，完美的答案是不存在的，但那些富有经验和创造性的程序员会通过一些分析和实验，找到一种看起来很不错的报价更新流程。

一个软件程序的效率不仅仅取决于程序员所选择的算法以及他所编写的程序的质量，编写软件的程序语言也会影响到程序效率。程序语言是一系列与英语比较类似的代码（编写软件也称为写代码），程序员需要按照严格的语义和语法规则使用各种指令和关键词向计算机输入代码，代码通过编译和链接最终被转换成二进制形式（0 和 1）的代码，它是计算机本身能够理解的最终指令形式。在高频交易软件领域，主流的编程语言有三种。C++ 语言是 C 语言的面向对象扩展版本^①，它继承了 C 语言“接近硬件”的特性，因此具有很高的效率。使用 C++，程序员可以编写出非常定制化的代码，这样的代码能够最好地满足软件对效率的终极需求。除了 C++，C# 和 Java 是另外两种流行的编程序言，它们也都具有面向对象的特性，虽然灵活性不如 C++，但是使用它们编程更加容易。

在其他因素相同的情况下，程序员用 C# 或者 Java 编写程序一般来说会比用 C++ 更快，并且程序也更不容易出错。但是使用 C# 或者 Java 编写高频交易软件有两个值得注意的缺陷。第一，正如我们提

① 广受尊敬的 C 语言是针对人们所说的面向过程编程而发明出来的，面向过程编程中，功能性的指令（做这个、做那个）和指令所操作的数据（对这个做、对那个做）被分开了。而 C++ 所支持的面向对象编程允许程序员将函数指令和数据捆绑在一起，称为对象。一个对象（类的一个实例）将它的的部分和函数部分封装（或者说隐藏）在一个良好定义的接口之后。面向对象概念由于可扩展性（例如通过对程序的修改向其中添加新的功能）和健壮性（例如，通过把可能产生的漏洞隐藏在对象内部而减少它对整个程序的危害性）而在程序员中广受欢迎。

到过的，相对这两种语言，C++ 让程序员能够编写出更加定制化的指令，包括指定特定的内存区域（使用*指针*，甚至*指针的指针*），甚至是操纵内存中的某些特定字节位置，而 C# 和 Java 没有提供这种让代码更“接近机器”的能力。这并不一定是件坏事，这个特性也使 C++ 程序出现漏洞的可能性要远远大于 C# 和 Java 程序，但它毕竟限制了程序员编写程序的定制性。第二个缺陷主要是关于内存管理的，说得更具体一些，是关于在程序运行完毕之后的内存释放，通常这件麻烦的事情称为*垃圾回收*。

使用 C# 和 Java 编程的一个吸引人的地方在于它们能够替程序员完成“垃圾回收”的任务，当程序发现一些内存不会再使用时，它就会自动将其清空，为其他任务腾出空间。而使用 C++ 编程时，程序员需要自己编写一些专用的代码来控制内存的使用和释放，如果不这么做，程序就会给一部分被使用的内存区域进行锁定，被锁定之后，这些内存即使在程序已经运行完成之后也不能被其他程序使用（这样的漏洞称为*内存泄漏*）。但自动的垃圾清理机制的问题在于，你基本上无法控制这些垃圾清理过程，而它们有可能在一些非常不合时宜的时候发生。举个例子，可能你的电子眼刚刚监控到了一个可能带来丰厚盈利的交易机会，而系统也正准备发出一张 IOC 委托单将它拿下，但它却突然发现自己因为正在清理垃圾而处于等待模式。这种不可控因素有可能决定你是否能够拿到一个交易机会。因为这样的一些原因，在构建最为高效和可控的高频交易系统组件时，C++ 可以说是最好的编程语言。不过 C# 和 Java 仍然是很流行的，而且也的确更适合用来构建高频交易系统中那些对时间效率要求不那么苛刻的组件（例如 GUI apps[⊖]和报告生成器等）。

⊖ GUI app = graphical user interface，即用户图形界面。

事实上，你对于操作系统的选择也会对软件的效率和可控性有影响。大多数高频交易系统都运行在标准的 Linux 或者 Windows 操作系统上。[⊖]现在的操作系统要处理的任务，已经远远不只是运行几个特定的程序，实际上它有一大堆麻烦而琐碎的任务需要处理，而大多数程序员甚至不知道这些任务的存在，这最终会导致即使写出最为高效的 C++ 程序，也不能保证计算机能够像你期待的那样完成任务。操作系统的处理能力需要在各种目的和需求之间进行分配，并且这是不受你控制的。的确有一些操作系统能够让你事先知道计算机将在什么时候处理你的任务，它们被称为“实时操作系统”（其中的大多数都是 Linux 的变种），在对效率和可控性有极高需求时，某些最苛刻、最激进的高频交易公司的确会考虑使用它们。

高频交易系统背后的代码量是非常庞大的，一个高频交易系统可能包含数十万甚至数百万条代码（在使用 C++ 编程的系统中，代码量更大，而使用 C# 或者 Java 编程的系统中的代码量会少一些）。每一行代码都可能对系统的质量优劣产生影响。同样，任何一行代码都是某种算法中的一步，而高频交易系统背后可能有数万个算法，其中每一个算法的选择都会对系统效率高低产生影响。毫无疑问，高频交易系统的开发者需要工作很长时间才能完成它（当然，他们也领着很高的薪水）。

出色的通信机制

在由分布式组件构成的系统的架构设计中，还有另一个非常重

⊖ 我不想在这里陷入关于 Windows 和 Linux 之间孰优孰劣的争论。两者都是很不错的操作系统，也各自具有相对的优缺点，选择哪一个都不能说更好或者更差。不过那些软件开发领域中最执著的理想主义者总会想着让两者比个高下。

要的方面，就是组件之间如何相互通信。每一个高频交易系统的设计者都必须认真地考虑这个问题。在组件之间传递的数据称为信息 (message)，而这些负责信息交互的机制则被称为系统的通信基础设施。信息在某些时候也称作事件 (event)，高频交易系统和其他一些与事件的交互联系紧密的系统也称作事件驱动 (event-driven) 系统。在这类系统中，任务并没有清晰定义的开端和结束。组件初始化之后就会一直处于等待状态，直到某些特定的事件发生（或者说某些特定事件的到达），然后通过一些确定的机制对其做出反应，并触发一些会引发其他组件反应的事件，系统就这样循环往复地运行下去。

“发布与订阅” (pub-sub) 是一个非常流行的通信机制。在这个机制中，组件被定义成消息的发布者、消息的订阅者，或者同时是消息的发布者和订阅者。例如，一个市场行情数据组件会负责把从外部接收到的实时行情在系统中发布；而另一些组件，比如定价引擎组件，则会是这些实时行情的订阅者。为了实现这个机制，系统通常会维护某种形式的消息类型目录，有时也被称作标签 (tag) 或者话题 (topic)，组件必须在发布或订阅消息之前进行登记（一个单独的组件通常不会既是一个标签的发布者，又是同一标签的订阅者）。

一个很容易想到的类比是杂志。出版商发布各种各样标题的杂志，而家庭客户则会订阅他们感兴趣的标题，当新的杂志出版的时候，它们被各自投递给不同的订阅者（一个杂志出版商同样会订阅其他的杂志来对市场上的竞争形势保持关注）。另一个类比则来自于微生物领域，不同的细胞会通过细胞膜分泌不同的化学物质，而细胞膜也允许一些特定的化学物质通过它而进入细胞。这种机制让细胞能够只得到它们想要的物质，而不论其他的细胞分泌了多少种不

同的东西。这很像一个通过“发布和订阅”机制来进行通信的系统内部所发生的事情，就像我们在图 4-1 中描述的那样，一些类型的消息被允许进入某个组件，而其他的消息则进不去。

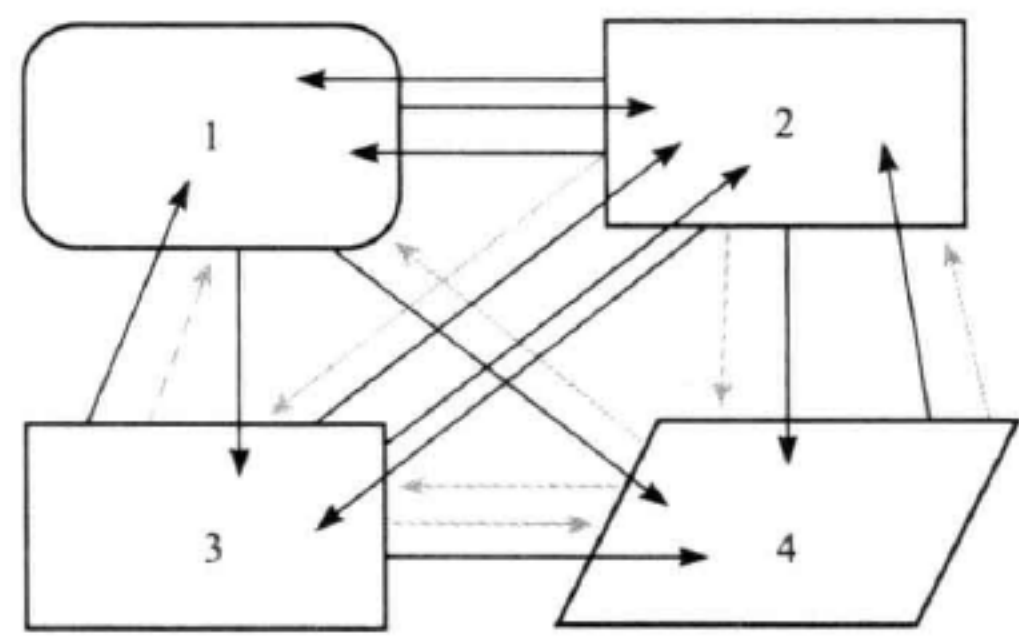


图 4-1 通信

这种“发布与订阅”的模式得到了广泛的接受和应用，不过实现这种机制则有多种不同的方法。在这里，我们恰好找到了一个例子能够证明，事实上不少高频交易系统的需求都能够通过第三方软件商提供的解决方案得到很好的满足。当然，一些高频交易公司总是想要自己构建系统的通信基础设施，但有一些很成功的公司通过把这项工作交给软件提供商，也很好地完成了这一部分的任务。这个领域中最著名的一个例子是由 Tibco 公司提供的叫做 Rendezvous 的产品。这个名声显赫的产品已经存在很多年了，而且我认为基本上可以说至少有一半的高频交易公司的系统核心中都能看到 Tibco 公司产品的影子。但 Tibco 公司在这个行业中还算不上垄断者，比它更年轻的公司，例如 29West 和 Tervela（我们会再次提到这家公司）都在这个领域向 Tibco 发起了挑战。

而在系统通信中，需要做出的另外一个重要选择是消息应该如何从一个组件发送到另一个组件。一种方法是在两个组件之间建立一个直接且无中断的连接，即 *TCP/IP* 套接字。这就像是和另外一个人的

直接电话线连接一样，这种电话连线只对你和对方开放，并且总是处于畅通状态。这种排他性，正是信道连接被认为是两个组件之间最可靠的通信方式的重要原因之一。但它天生就只是一种一对一的解决方案，当成百上千个组件需要互相通信时，在每一对可能发生信息传递的组件之间都建立套接字显然是一种笨拙得可怕的办法。

组播（multicast）是一种流行的替代方案。在这种方式下，一个信息发布组件通过网络将消息经由广播的方式发布出去，这个消息并没有特定的订阅者，它仅仅是像无线电天线一样把消息扩散出去，而是否接收消息则是那些可能需要订阅这个消息的组件的任务了。如果一个组件想要接收一个组播信息，它必须使用接收天线来对消息进行监听（网络中的路由器必须经过一定的设置来保证组播信息能够最快地到达接收者）。组播是一种很实用的同时向多个订阅者发送消息的方法，但它并非一种传统的可靠性很强的消息传递机制。[⊖]组播无法提供 TCP/IP 套接字的可靠性来保证某条消息能够一次就发送到其订阅组件，就像通讯公司无法保证你拨出的第一个电话就能拨通一样。不过 TCP/IP 套接字和组播并不是仅有的通信机制选项，高频交易系统的设计师也的确会花上不少时间在不同的选项之间进行权衡，试图选出最合适的通信机制。

内核之外的任务处理

在计算机技术的发展史上有一个很有意思的现象，从计算或处理任务发生的具体位置来看，这些任务在逐渐分散化。我们所谈论的现代意义的计算机，与早期的一些计算设备之间存在两个主要的差

[⊖] 在这个领域中，现在已经出现了一些更可靠的多重广播协议。

别，而它们也被视为 20 世纪中叶以来计算技术爆炸式发展的最重要原因。第一个差别是可存储式程序的概念，可存储式程序指令并不是被固化在物理电路中的，而是保存在一些外部介质中——早期的穿孔卡片、纸带和当今的磁盘、闪存等——以备计算机在需要的时候调用。第二个差别则是通用计算机的概念，这主要依赖于一个能够处理任何能够用可存储式程序指令解决的任务的中央处理器。任何买过电脑的人都知道，市场上出售的处理器计算能力在不断增强。在奔腾（Pentium）系列处理器出现之前，英特尔 486 是处理器世界中的王者，而现在，奔腾的地位也被至强（Xeon）系列和其他的一些处理器取代了。你不可能总是赶上处理器更新换代的速度，而那些电脑制造商则对这一点很欢迎，因为处理器的不断更新让我们每两三年就得换一次电脑了。

实际上要制造出下一代让人心动的计算机，关键就在于如何让处理器拥有更强的计算能力，作为整台机器的核心，它能够完成你交给它的几乎所有任务。但是有些计算机的用户在经过更深入的思考之后，却得出了一个一眼看上去似乎是南辕北辙的想法。他们提出了这样一些问题：把所有计算任务的处理都压缩到一块小小的硅片上真的是一个必要或者说聪明的做法吗？如果你需要一块电脑芯片来完成某种任务，真的需要设计一块芯片来完成所有不同的任务吗？难道我们就不能至少把一些任务从 CPU 的内核转移到别的地方吗？最后，难道就不会有一些计算机应用，例如高频交易，使用 CPU 之外的硬件会更好吗？对于最后一个问题，答案似乎的确是肯定的。

这种内核之外的任务处理的一个例子是计算一个计算机网络中的用户（基本上是所有的用户）有哪些很享受无知的感觉。如果你

怀疑这样的用户是否真的存在，你可以去问一些你认识的非程序员用户：“你喜欢你 Windows 笔记本上的 TCP/IP 栈吗？”然后看看有多少人会用不高兴的眼神看着你。你可以接着问：“你知道，其实我指的是 Winsock！”如果你不介意对方是否认为你已经疯了，你还可以继续这样的问题。想象这样一项任务：运输一些小型产品（例如办公用品）穿越全国，假设每种产品（订书机、橡皮、钢笔和其他的东西）都装在贴着各自标签的箱子里，这些箱子装在小卡车里，这些小卡车又装在大拖车里，大拖车放在平板列车上。现在火车到了你所在的站台，而你要取下那些 2 号铅笔[⊖]，你该怎么做呢？首先，你需要找到那一节列车车厢，然后找到那辆正确的拖车，然后是装着铅笔的小卡车，然后是装着 2 号铅笔的箱子，最后取出那些你需要的 2 号铅笔。如果你想再把这些铅笔放回去，那么就得有人按着相反的顺序把这些事情再做一遍。

我们刚才所举的例子并不能和内核外的任务处理形成一种完美的类比关系（一个更好的类比可能是把铅笔不断地拆分，先拆分成分子，然后是原子），这种分层的传递机制实际上很像数据在网络中的传递过程。在每一层的节点上，都需要有硬件来完成打包和拆卸的工作，这种硬件通常都是 CPU。一般来说，当一个数据包到达某个计算机节点，节点的 CPU 必须停下它正在做的事情，然后打开数据包，把它的内容放进内存，从而让程序能够访问其中的数据。而当经过处理后的数据包需要重新回到网络中，CPU 会再次中断它正在处理的任務，把数据从内存中取出，打包之后发回到网络中。

对于 CPU 来说这实在太累了！难道就没有一些别的硬件能够替

⊖ 我儿子曾经问我：“爸爸，既然 2 号铅笔这么受欢迎，为什么人们不干脆叫它们 1 号铅笔呢？”我问他：“你是自己想出这个问题的还是从别人那里听到这个问题的？”我儿子说：“从别人那儿听来的。”我说：“那就是了。”

它完成这类任务吗？答案是，有。一种叫做 TOE 卡（*TCP/IP* 减负引擎）的硬件能够完成在网络和内存之间传递数据的功能，从而把 CPU 从这种任务中解放出来专心去做其他的事情。由于每台处于网络中的电脑实际上都在永不停歇地和网络进行数据的交互，TOE 卡的出现就从处理器上分担下一块很大的任务量，从而让处理器能够把更多的时间放在识别交易机会、下单去试图从交易中盈利等更重要的任务上。

实际上，任何在传统上由 CPU 来完成的计算任务都是内核之外任务处理的备选项。我们之前提到的 Tervela 公司提供的 TMX 信息交换机服务就是一个例子，它使用内核外的硬件实现了一个完整的“发布与订阅”信道。这种服务很像一个基于软件的“发布与订阅”机制，其中的组件会发布或者订阅它所感兴趣的数据，但实际数据的提供和存储都是由硬件上的物理电路完成的，而并不是内存。TMX 的数据吞吐量非常惊人，它为信息的发布者和订阅者提供了一种中介服务，正因如此，一些工程师开始争论中介服务的固有延迟是否会使得一个基于硬件的信息交换机比直接的软件连接更加高效。目前，TMX 服务被高频交易业内的很大一部分公司所采用，并且已经成为通用计算任务下放的一个重要范例。

还有另一个比 TOE 卡和硬件信息交换机更重要的内核外任务处理的例子，它不仅仅是把计算任务放到其他一些专门用途的设备上，还把计算任务放到一个为通用计算任务重新设计过的专用设备上——这个设备就是图像处理器 GPU，这种为通用计算重新设计 GPU 的概念称为 GPGPU（general-purpose computing on a graphical processing unit）。有意思的是，在图像处理器最初出现时，其目的就是把图像处理的计算任务从 CPU 上转移出来。就像在网络和内存之

间传递数据一样，图像处理的基本任务也是非常简单的——确定电脑屏幕上一个像素点的颜色和强度，但是这种基本任务的数量却多得数不清。GPU 的出现让 CPU 从这种繁重而单调的工作中解放出来了。而在 GPGPU 的概念下，GPU 所处理的任务甚至可以和图像处理一点关系都没有。

现在我们可以确定，一个 GPU 没办法完成 CPU 所做的那些复杂的计算，它只能够处理简单的任务，但它能在很短的时间内并行处理大量的简单任务。正因如此，高频交易中的许多计算并不适合使用 GPGPU 来处理，但是在交易过程背后却有许多工作非常适合交给 GPGPU，其中的一项任务就是使用一种叫做蒙特卡洛模拟的方法为期权定价。众所周知，期权的价格是由其标的证券（例如股票）价格衍生而来的，而衍生证券的价格被假设为一随机路径。[⊖] 使用蒙特卡洛模拟法，建模者可以生成（或者说模拟）大量的可能的价格路径，在每一条路径上确定期权的价值，并且通过将这些价值进行平均，得到一个比较合理的期权价值估计。计算一条随机路径上的股票价格是一个相对来说简单的任务：给定一个初始价格，在这个价格之上加上或者减去一个随机数值，从而得到下一个时点的价值，然后重复这个过程。对于 GPGPU 来说，这样的任务是小菜一碟。

并行处理

我们已经看到，在 GPGPU 的帮助下，一个大问题（计算期权价格）是如何通过分解成大量的小问题，并同时处理这些小任务而完成的。这样的并行处理并非仅限于 GPGPU。我们之前所讨论过的分

⊖ 准确地说，价格过程服从的是几何布朗运动。

布式和负载平衡问题中已经提到过它，在这个问题上，高频交易系统的设计者需要确定在一台机器上需要为某个组件复制的最优实例数量，以及运行该组件的最优机器数目。即使在一个独立的 CPU 上同样存在着并行计算的机会。例如，一个四核处理器实际上就非常类似于四个被捆绑在一起的处理器，它能够让程序员编写最多能同时处理四个任务的程序。而 GPGPU 能够并行处理的任务就远远不止四个了，它可以同时处理成百上千个类似的任务。当然，CPU 的生产商也显然不会满足于四核、八核，甚至是十六核的处理，在不远的将来，拥有数百个核的 CPU 的出现并非是不可想象的，那时候，CPU 也许将拥有和今天不可同日而语的并行处理能力。

直连交易所接收数据和执行指令

高频交易公司需要能够在低延迟的条件下访问交易所，这主要是因为高频交易中的两项重要任务要求在低延迟下完成：第一项任务是市场数据的接收，第二项任务是提交委托单或报价（这项任务也称为交易执行）。由于延迟很自然地会把速度降下来，所以那些最激进的公司都会用尽可能直接的方式接入交易所。

市场数据主要包含委托单簿的最新变动（新的申买和申卖）、交易执行完成通知（某个买单和某个卖单被撮合）、市场的当前状态（闭市、交易中、中止等）等。交易者可能会直接从交易所获得市场数据，也可以通过统一数据平台间接获得转发的数据，这取决于所交易证券的类型。例如，基本上所有的股票市场行情数据都可以从统一证券行情协会（Consolidated Tape Authority, CTA）获得。目前，CTA 是纽约－泛欧交易集团（NYSE Euronext）的一部分，它

提供两种数据馈送。一种是由统一报价汇总系统（consolidated quotation system, CQS）发送的委托单簿变动信息，另一种是由统一证券行情系统（consolidated trade system, CTS）发送的交易信息。参与交易所会将实时的市场数据流发送给 CTA，CTA 对此进行汇总，再直接发送给数据接收者，或者发送给向其客户转售数据的数据提供商。

大多数股票交易市场也直接提供数据，而期权市场的数据同样可以通过直接通道和统一转发通道两种方式获取。期权市场的统一转发数据由期权报价处（options reporting authority, OPRA）提供^①，它包括所有向其传送数据的期权交易所的报价、交易和市场状态等数据。^②期货市场的数据则是由芝加哥商品交易所和洲际交易所直接提供。^③

除了 CTS/CQS 和 OPRA 的数据之外，最激进的高频交易公司会从他们进行交易的每一个交易所直接获取市场数据，他们不使用第三方数据提供商出售的数据以避免由中介转发而自然而然导致的延迟。^④你也许会好奇，既然可以直接从交易所接收，为什么大家还要从统一平台接收数据呢？这是因为直接接收的方式天然的有更高

① 似乎高频交易业内有一半人都会把 OPRA 这个缩写词读成著名脱口秀主持人 Winfrey 女士的名字，而另一半人则会将它读成“opera”（歌剧），这里我指的是罗西尼、瓦格纳、莫扎特这些音乐家创作的那种艺术形式。

② 截至 2010 年 6 月，向 OPRA 发送数据的交易所包括芝加哥期权交易所（CBOE）、国际证券交易所（ISE）、纽交所群岛交易所（NYSE Arca）、纽交所全美证券交易所（NYSE Amex）、纳斯达克期权市场、纳斯达克费城股票交易所、波士顿交易所（BOX）和 BATS 交易所。

③ 这些期货市场数据主要是由一种叫做“安全金融交易基础设施”（Secure Financial Transaction Infrastructure, SFTI）的设备进行发布的，它是一种高可靠性、低延迟的专门用于发布市场数据的计算机网络。和 CTA 和 OPRA 系统一样，它也由纽交所技术公司所有和运营。

④ 一些公司也可能很明智地使用由专门进行市场数据处理软件研发的公司所开发的产品，当然，它们不会直接使用软件提供的数据，而是在自己的数据中心中将软件完整地实例化。

的成本，因为它需要维护与多个交易所的连接而不仅仅是与统一平台的单一连接。对大型的高频交易公司而言，成本并不是什么问题，但他们仍然会使用统一平台。首先，这是因为没有哪种市场数据的馈送方式是完美无缺的，拥有多个数据源能够帮助交易者更好地判断例如一个异常的价格跳动的真实性之类的问题，因为它可以对不同来源的数据进行比较。另外，也有可能在某些情况下，使用统一平台的数据可以比直接接收更快地捕捉到价格变动。这与交易所如何发送价格变动的信号有关：一般来说，交易所会在价格变动之后立即将这一信号传送给 OPRA 之类的统一数据转发平台。但是它们在向直连数据订阅者发送信号时，可能会使用一种轮询策略，例如每隔 500 微秒发送一次。如果恰好在一个轮询之后发生了价格变动，从 OPRA 订阅数据的用户可能会比直接接收交易所数据的用户更快地获得这个价格变动的信号。

在与交易所建立交易执行的连接时，高频交易公司会面临与建立市场数据连接时类似的选择。每家交易所都会允许交易公司用直接的方式下单，例如使用专用的数据线路。同样，许多交易经纪商也会接受客户的委托单并将它传递到交易所。[Ⓐ]就像你预期的那样，高频交易公司希望减少延迟，所以他们会选择直接向交易所发送其委托单和报价，而不经任何的交易中介。[Ⓑ]

原生 API

一旦高频交易公司决定不经过任何中介，与交易所直接建立连接，它就必须做出另一个重要的抉择。这里所面临的选项有两

Ⓐ 高盛公司提供的 REDI 系统在这个领域中应用广泛。

Ⓑ 与交易所之间建立直连途径导致了“裸接入”的兴起，裸接入是一种广受争议的连接方式。我们在之后讨论高频交易的益处与风险时会再次提到这个概念。

个：第一个是使用业界标准的通信格式来接收市场数据、下单并监听交易回报等，这种格式被称为 *FIX*，其全称是金融信息交换协议（financial information exchange protocol）；另一个选项则是通过每个交易所特有的原生应用程序接口（application programming interface, API）编写能被交易所电脑直接解释的代码。

FIX 协议的优势是显而易见的。不论你与哪家交易所进行通信，你都能够使用完全相同的方式编写你的指令（顺便说一下，我们在这里所讨论的是非常具体的代码编写方式，这是程序员才需要考虑的问题，例如，我应该把交易价格放在数组的第四个还是第五个位置上？我应该在哪儿指定我所要交易的合约名称？诸如此类）。这种通用性能够为高频交易公司省下一大笔软件开发的费用。如果公司使用原生 API 与交易所通信，一旦需要与一家新的交易所建立连接，程序员就必须从零开始编写程序。而使用原生 API 的程序进行规划、编码和测试，常常需要花掉好几个星期，甚至是几个月的时间。

那么为什么会有人不使用 *FIX* 协议呢？这与程序的翻译时间有关：当交易所接到一条原生格式的消息时，并不需要进行什么翻译。举个例子，一条原生格式的高频交易委托单在接收到之后会立即被处理。当交易所接收到一条 *FIX* 格式的消息时，则需要首先把消息从 *FIX* 格式翻译成原生格式，然后才会被处理。

我们在这里说的时间并不是很长，在大多数情况下都远远小于 1 毫秒。但是，那些最激进的高频交易公司愿意进行这样的软件开发投入，哪怕仅仅是为了让他们的委托单再快那么一丁点儿。

值得一提的是，一些交易所不允许使用原生 API 进行连接。他们可能会只提供 *FIX* 格式^①的市场数据，或者仅仅接受 *FIX* 格式的委托

① 一种叫做快速 *FIX* 的 *FIX* 协议变种，目前广泛应用于交易所的数据发送。

单。在这种条件下，由于没有人会在 API 上拥有优势，高频交易公司显然也很乐于使用 FIX 协议。

共置

所有高频交易公司在追求最低时间延迟时都无法回避的一个问题就是其高频交易系统服务器的物理位置与交易撮合引擎之间的距离。在其他条件完全相同的情况下，系统服务器与交易所的市场数据服务器或交易撮合引擎最近的那家公司相比其他公司将获得明显的优势。考虑两家高频交易公司 A 和 B，它们都监测一只在纽交所群岛交易所交易的股票，股票的价格在 1.00 ~ 1.10 美元之间变动，两家公司都准备在股价达到了 1.09 美元时就下单交易。A 公司的电子眼位于新泽西州的卡特里特——与纳斯达克撮合引擎位于同一座建筑物中，而 B 公司的电子眼则位于新泽西州纽瓦克的一间数据中心。假如这种情形真的发生了，比如一位投资者以 1.09 美元的价格申报卖出 500 股，两家公司同时探测到这个信号，并几乎同时发出了委托单（我们假设它们同时收到市场数据的变化）。A 公司的委托单到达了撮合引擎并且完成了交易，B 公司的委托单晚了 300 毫秒到达，什么也没得到。没有人希望自己在这种情况下处于 B 公司的位置。因此，在高频交易的世界中，共置几乎是必需的。

每家股票交易所、期权交易所或者期货交易所都拥有唯一的一台主撮合引擎（见表 4-1）。这台主撮合引擎位于一家属于交易所或者从托管服务提供商（例如，位于新泽西州斯考克斯的 Equinix，还有位于新泽西州威霍肯的 Savvis）租用来的数据中心。托管服务提供商的作用仅仅是出租机架空间，将服务器安装在这里，并且使用交叉连接把服务器与交易所撮合引擎连在一起。你可以只租用一个厚度

为 2U（8.9 厘米）的机架空间，它足够装下一个大比萨饼盒大小的服务器。你也可以租下一整个机架的空间，甚至是好几个机架的空间。当然，这些服务器不仅与交易所撮合引擎相连，同时也与公司高频交易系统中的其他服务器相连，并最终被连接到公司的交易台，使交易员能够实时监控系统的行为。大型高频交易公司不会把所有的高频交易系统硬件都放在一个地方，而是把它们分散在一个广域网（wide area network，WAN）中的不同地点。那些在四大股票交易所和所有的期权期货交易所都进行交易的公司会在表 4-1 中的每一个地点都部署自己的系统。

表 4-1 交易所撮合引擎所在位置

交易所	撮合引擎
股票	
BATS	新泽西州威霍肯
Direct Edge	新泽西州泽西城（2010：新泽西州斯考克斯）
纳斯达克（NASDAQ）	新泽西州卡特里特
纽约证券交易所	新泽西州威霍肯（2010：新泽西州莫瓦）
期权	
波士顿期权交易所（BOX）	新泽西州纽瓦克（2010：新泽西州斯考克斯）
芝加哥期权交易所（CBOE）	伊利诺伊州芝加哥
CBOE C2（2010）	新泽西州斯考克斯
芝加哥商品交易所（CME）	伊利诺伊州芝加哥
洲际交易所（ICE）	伊利诺伊州芝加哥
国际证券交易所（ISE）	泽西城（2010：新泽西州斯考克斯）
纳斯达克期权市场（NOM）	新泽西州卡特里特
纳斯达克费城期权交易所（NASDAQ PHLX）	新泽西州卡特里特
纽交所全美期权交易所（NYSE AMEX）	新泽西州威霍肯（2010：新泽西州莫瓦）
纽交所群岛期权交易所（NYSE Arca）	新泽西州威霍肯（2010：新泽西州莫瓦）
期货	
芝加哥商品交易所（CME）	伊利诺伊州芝加哥
洲际交易所（ICE）	伊利诺伊州芝加哥

交易设施的共置到底会在多大程度上影响交易速度呢？假设一家芝加哥的高频交易公司在纽约的交易所从事交易（反过来也一样），把交易设施放在纽约将能够节省 12 ~ 15 毫秒的时间，这大概也就是数据在这两个城市之间传递的时间，它主要取决于光缆中信号的速度；而在纽约市的区域内，数据包从一个主共置点传到另一个主共置点大约需要 300 ~ 500 微秒。除了地理距离之外，数据链路的直接性也会影响实际的延时效果。从 A 点到 B 点之间的网络连接很少是一条完整的光缆，它可能是由许多条光缆连接而成的，这些光缆由路由器或是交换机连接在一起，就像我们所知道的，两座相距遥远的城市之间的通路一般不是一条单独的道路，而是有许多更短的道路通过一些中间结点相连而成。在网络中，每一个这样的节点（或者称为“一跳”）都可能导致几微秒的延迟。

高速数据通道

最激进的高频交易公司会在全美所有的交易撮合引擎所在的位置部署交易服务器，通过将高频交易系统分散到多个地点，力争最小化与各交易所之间的通信时间。但这也会导致一些小摩擦。因为一家公司的许多相同或者相似的产品会在多个地点运行，让它们的行为协调一致就变得非常关键。举例来说，假设一家公司在威霍肯的电子眼监控到了在纽交所交易的某只股票上一个大的买入单，交易程序决定吃掉其中的一部分交易量，所以它会发出一个下单指令并完成，而这家公司位于卡特里特的电子眼此时就需要尽快收到这个信息了，否则它也会做出一个同样的决策并在纳斯达克市场完成同样的交易，这会导致公司持有比预期更多的头寸。这就像你派出了

一整队旅行销售商，他们在不知情的情况下销售了比你的公司所拥有的更多的产品，而这正是由于他们之间缺乏足够快速的沟通机制。

在一个完美的高频交易系统中，每一个交易服务器都应该在进行交易之后让其他的交易服务器即时得到这个消息。当然，从理论上说是不可能实现的，因为即使是光的传播也是需要时间的。但是仍然可以通过一定的代价让高频交易系统中的节点能够以最小的延时进行通信。一些高频交易公司也因此投入了大量的研究和规划来建立属于他们自己的广域网。

理想的高频交易广域网需要让其中的节点以足够的带宽互联，从而消息能够从一个节点以物理定律所允许的最快速度到达另一个节点。带宽衡量的是多少单位（通常是比特）的数据能够在一定单位（通常是秒）时间内通过一个网段。例如，一个带宽为 1 Mbps 的网段能在 1 秒之内大约平均传送 100 万个比特，一个带宽为 1 Gbps 的网段能够在一秒之内传送 10 亿个比特。今天^①，那些激进的高频交易公司会将他们自己网络中的节点以几十到几百 Mbps 的带宽相连接，而这个数字的大小主要取决于他们要交易的证券数量。网络工程师需要对数据通路“量体裁衣”，也就是为两个网络节点之间的通路确定合适的带宽，这并不是一件容易的事情，其中的困难主要在于数据并不总是在节点之间平稳地流动，数据流常常在随机的时间段内“爆发”，这取决于市场活动的频繁程度，而估计这种数据流“爆发”的规模和持续的时长既需要一些概率计算，也需要经验。对这种“爆发”的高估会导致公司在所有连接上都以 Gbps 级的带宽规模进行配置，这即使对于最有钱的高频交易公司，这也是非常可观的开销；而对这种“爆发”的低估则会导致网络中的数据阻塞。

① 这里指 2010 年年初。

120 打开高频交易的黑箱

All About High-Frequency Trading

假设有一条小型高速公路，它在正常情况下能承载每分钟 100 辆车的“带宽”。如果在这条高速公路边上有一个篮球场，当一场比赛结束时，很可能会有 500 辆车同时需要使用这条高速公路。无论如何，这些车最终都会通过这条高速公路，但是相比于一条被设计为承载每秒 500 辆车流量的高速公路，速度就会慢得多了。同样的事情也会在数据流“爆发”超出带宽时发生。所有的消息都被阻塞并处于等待状态。在高速公路和篮球场的例子中，偶尔的阻塞也许是能够接受的。但是在高频交易中，这会导致某些节点在它们必须对市场变化做出反应的时候，可能根本收不到这些信息。

要将所有位置的节点连接在一起有许多方法，它们的成本各不相同。比较低端的解决方案是*虚拟专用网络*（virtual private network, VPN），有时也称为*网络隧道*。在这种情况下，你的数据实际上是在公用的互联网上传递（至少其中的一段是在互联网上），但它们被封装在受保护的数据包中。这就像你雇用了一个保安运输公司，把你的东西放在不被人注意的、拥有武装保护的货车内通过公用的高速公路运往新家。你的东西是受到保护的，但是当交通压力很大时，运输过程还是会受到延迟的影响。

对于高频交易广域网而言，一种非常有吸引力的低延迟数据通信方式是 *SFTI 网络*（全称是“安全金融交易基础设施”，secure financial transaction infrastructure），它由纽交所技术公司所有并运营。你可以将 SFTI 想象成一个巨大的拥有惊人带宽的光纤环，它从美国的各大城市穿入穿出，就像美国的州际高速公路一样。[⊖] 每一个股票交易所、期货交易所和期权交易所都用一个超高带宽的数据链路将

⊖ 实际上，SFTI 包含了两个这样的光纤环，它们传递着一样的数据，这是为了确保在其中一个出故障时，另一个仍然能够正常使用。

它的撮合引擎以及市场数据发布端口连接到 SFTI 的光纤环上。现在，如果你的高频交易公司想要连接到这些交易所中的任意一个，你只需要与通过 SFTI 光纤环上的许多专为获取数据准备的“匝道”之一建立连接就行了。即使与 SFTI 建立了连接，还是需要为你实际使用的带宽和交易所连接付费，但 SFTI 为你提供了一种非常简便的让你的高频交易“广域网”运行起来的方式。

另一种实现你的“广域网”的办法是使用一个由私人数据通信提供商拥有和提供的被称为外联网的服务。在这种方式中，数据通信提供商建立了私有的“数据高速公路”网络，将包括交易所和高频交易公司在内的许多节点连接起来，而你则为使用这种网络进行数据通信付费。你仍然在与别人共享这些“高速公路”网络，但是这种“外联网”的所有者则会控制在这些“高速公路”上同时行驶的“汽车”数量，所以你所使用的带宽还是能够得到比较好的保证的。

而在成本最高的那端，则是使用你自己私有的专用数据通路——只有你能够使用它。你可能需要为此付上一笔数目不小的钱，但能够确定这条数据通路在你需要它的时候一定是畅通无阻的。即使你的公司已经在专用的高带宽数据通路的玩家俱乐部里了，你还是在选择的时候尽量小心一些。举个例子，两家提供商向你推荐的数据通路可能看起来没有太大的差异，但可能其中的一个比另一个多绕了一点弯，这就会增加几跳，并会带来更多的延迟。

实时监控

在构建一个高频交易系统的过程中你还需要注意另外一件重要的事情。也许你的系统看起来已经很好了，但如果在这里你做得不

122 打开高频交易的黑箱

All About High-Frequency Trading

好，很有可能会把整个系统搞砸。由于所要处理的事情的高度复杂性，高频交易系统天生就具有很复杂的结构：多个交易所、有着不同相关性的数千种可交易资产、激烈的市场竞争、复杂的监管限制，还有许多其他的因素都导致了这种复杂性。如果你的系统能够更好地监控自己的行为并将信息反馈给那些需要知道这些信息的人，那么你所处的状况就更有利。即使是中小规模的高频交易系统也是由数千种不同的组件（计算机程序）组成的，它们运行在不同的地点，并且或多或少会彼此影响。那些坐在交易台上的交易员和系统支持工程师必须能够保证所有的一切都在正确的轨道上。当有些事情出错了（这实际上一直在发生），这些人需要以最快的速度知道到底发生了什么，以及需要做些什么来解决问题。看着系统做了一笔坏交易——卖得太低或者买得太高——感觉一定不会太好，而如果你不知道这是为什么，那么你的感觉就更差了。

具有讽刺意味的是，这种情况下的问题并不在于你拥有的信息太少，而是你拥有的信息太多。在很多例子中，几千个同时运行的进程都在持续地向日志文件写入正在发生的事情（例如，获得了一个新的市场数据、请求贴现率数据、接收贴现率数据、计算理论价格，等等）。这些日志文件在你意识不到的时候已经变得非常大了。最终，你需要用这些日志文件来诊断到底出现了什么问题，你会在一堆日志中的某个地方找到它。这些日志对交易员来说并没有什么用途，当计算机做了一大堆坏的交易时，交易员除了骂娘什么也做不了。他们仍然需要等着程序员过来，找到那个出了问题的日志文件，一行行的读，思考哪里出了问题，并与其他程序员进行讨论，以对整个状况做出判断。你知道，这样的方式很不好。高频交易系统的设计者可以用很多方法实现这种实时监控，但是并没有一种方法就

绝对比其他的方法好。无论采用哪种办法，这些事都还是必须做的。

很显然，本章讨论的高频交易系统实现策略并不是唯一的版本。那些在高频交易公司工作的聪明而富有创造力的人也一定会想方设法设计出新的策略来让自己在竞争中获得一些优势。但是本章的内容仍然能让你对于高频交易世界里的这个重要领域的一些基本理念有不错的了解。那么现在，让我们来看看一个采用了许多上面介绍过的理念实现的高频交易系统看起来到底是什么样子。

第 5 章

All About High-Frequency Trading

揭开面纱

现在，让我们以一个假想的系统为例，慢慢揭开高频交易系统的面纱，看看下面到底有些什么。我们在本章讨论的系统并不是基于任何一个实际中的高频交易系统，但是你会从它身上看出一个真实的系统大概是什么样子，以及它能复杂到什么程度。可以想象，真实高频交易系统的详细规划蓝图对于那些建造这些系统的公司而言，是需要高度保密的。

这个系统会差不多实现我们在之前所讨论过的每一个策略。这意味着它须能够交易美国证券市场这个大超市里几乎所有物品——股票、期权、指数期货，而我们对它的描述过程也将进一步说明这些市场是多么紧密地相互关联的。我们的系统将 与两家最大的期货交易所、所有的期权交易所以及美国四大证券交易所——纽交所、纳斯达克、Direct Edge 和 BATS 建立连接。如果要把这个范围扩大到其他一些信息公开的另类交易系统也并不是一件难事。[⊖]

⊖ 要把系统的交易范围扩大到暗池，从理论上说也是很容易的；但是，由于暗池中
没有公开的市场报价，在其中进行交易就要求系统对于报价引擎和电子眼的逻辑
进行一定的修改。

我们所描述的是一个前台系统，其功能主要是定价、交易，以及风险管理。我们将不会讨论关于清算、保证金以及会计之类的后台任务。毫无疑问，这些后台工作对于任何交易实务都非常重要，但是它们几乎不受交易速度的影响，因此不在我们的讨论范围之内。

在描述这个系统之前，首先假设我们能够直接接入各个交易所获得原始市场数据，并且有直接在每个交易所交易的权利。我们不使用任何经纪自营商或者其他中介来间接地与交易所的撮合引擎建立连接。实际上，每家交易所提供直连权限的方式都不完全相同。在某些交易所，你仍然需要一个真实的交易所会员资格来获取直连的权限（例如芝加哥期权交易所，在本书写作时，其相关规则正在修订中）；而在另一些交易所，你仅仅需要每月支付一个固定费用就能够达到目的，这些交易所会提供一个服务清单，你可以选择其中的服务并为之付费。此外，我们还假设系统的速度是符合各种法规和每个交易所的交易规则的，并且我们已经通过了各种必需的认证，例如交易员检查等。最后，我们还假设我们与一家结算公司达成协议，将由这家公司负责在每日交易结束后对我们的交易进行结算，并在这家公司放置了足够的保证金，或者我们已经采取了必要的步骤自己成为了一家清算公司。

高频交易服务器的位置

我们对系统描述的第一步从地理概念开始，我指的是高频交易服务器的实际物理位置。我们的服务器分别部署在 8 个不同地点的数据中心里，其中的 6 个共置在交易所附近的站点，目的是以低延迟访问交易所的撮合引擎和市场数据发布端口，另外两个则是控制服务器，它们的工作是负责协调各个远程服务器的行为。这两个控制服务器

还包含了交易台和后台支持的功能，在必要时，会有人负责通过它们对系统进行监控和维护。

显然，我们把服务器部署在交易所附近的目的是减小市场数据获取和交易执行的时间延迟。当交易所发布了一个新的数据行情，我们希望能够尽快地获得它；同样，我们也想在尽可能短的时间内获得 CTA 和 OPRA 这样的统一数据平台发布的行情，因此我们也会在它们的数据发布源附近部署服务器。当下单或报价时，我们希望指令能够以最快的速度到达交易所。我们一共要与 15 家交易所建立连接。幸运的是，其中的一些交易所的撮合服务器位于相同的数据中心，因此我们可以使用同一个服务器来处理与这几个交易所之间的连接，所以实际上我们所需要的用于连接交易所的服务器数目只有 6 个。表 5-1 给出了远程服务器的列表、每个服务器连接的交易所，以及它们在各交易所交易的产品类型。[⊖]

表 5-1 共置服务器站点所在位置

站点	地址	撮合引擎 / 市场数据源	产品
卡特里特	Verizon 机房， 新泽西州卡特里 特市联邦大道 1400 号	纳斯达克股票市场（NASDAQ） 纳斯达克期权市场（NOM） 纳斯达克费城期权交易所 （NASDAQ PHLX）	股票 期权 期权
芝加哥湖滨	Equinix 机房， 伊利诺伊州芝加 哥市中央大道 350 号	芝加哥商品交易所（CME） 洲际交易所（ICE）	期货、期货期权 （包括标普 500、纳 斯达克 100 和欧洲美元） 期货、期货期权 （罗素 2000）
芝加哥环线	CBOE 机 房， 伊利诺伊州芝加 哥市圣拉萨尔大 街 400 号	芝加哥期权交易所（CBOE）	期权

⊖ 截至 2010 年 3 月，表 5-1 中所提供的主撮合引擎以及托管机房位置的信息都是准确的。会有些不太确定的问题对表中信息的准确性造成影响，比如芝加哥期权交易所什么时候会上线 C2 系统等。但这张表基本上能让你知道各交易所的主撮合引擎分别在什么地方。

(续)

站点	地址	撮合引擎 / 市场数据源	产品
莫瓦 ^①	纽交所技术公司机房，新泽西州莫瓦市麦克阿瑟大道	纽约证券交易所 (NYSE) 纽交所全美期权交易所 (NYSE AMEX) 纽交所群岛期权交易所 (NYSE Arca) CTA/OPRA (市场数据)	股票 期权 期权 股票 / 期权统一市场数据
斯考克斯	Equinix 机房，新泽西州斯考克斯市斯考克斯路 755 号	国际证券交易所 (ISE) 波士顿期权交易所 (BOX) ^② Direct Edge CBOE C2	期权 期权 股票 期权
威霍肯	新泽西州威霍肯市港口大道 800 号	BATS 股票交易所 BATS 期权交易所	股票 期权

- ① 纽交所计划在 2010 年下半年启用位于莫瓦的新交易设施，在此之前，所有列出的位于莫瓦的设施实际上都是在威霍肯。
- ② 本书写作于 2010 年 3 月，在本书写作时，BOX 的交易设施还位于新泽西州纽瓦克哈尔赛大街 165 号的 Level-3 数据中心，他们最近宣布将会把交易设施搬到斯考克斯。

图 5-1 展示了我们的系统所运行于其中的广域网，它使用我们之前讨论过的“外联网”来实现。同样，我们可以使用 SFTI 来构建部分或全部的广域网，也可以为图中所有的节点都使用专用的数据通路，只不过这些方法更加昂贵。广域网所要做的主要事情是使得网络中的任何节点都能够相互通信。可以看到，我们的广域网中含有两个控制站点，一个在芝加哥，一个在新泽西，其主要功能是协调其他远程站点的行为，以及让交易员和负责系统支持的程序员对系统行为进行监控。还可以看到，我们的系统和交易所之间并不存在任何的中介，市场数据将直接从交易所传给系统，而不通过任何的第三方。

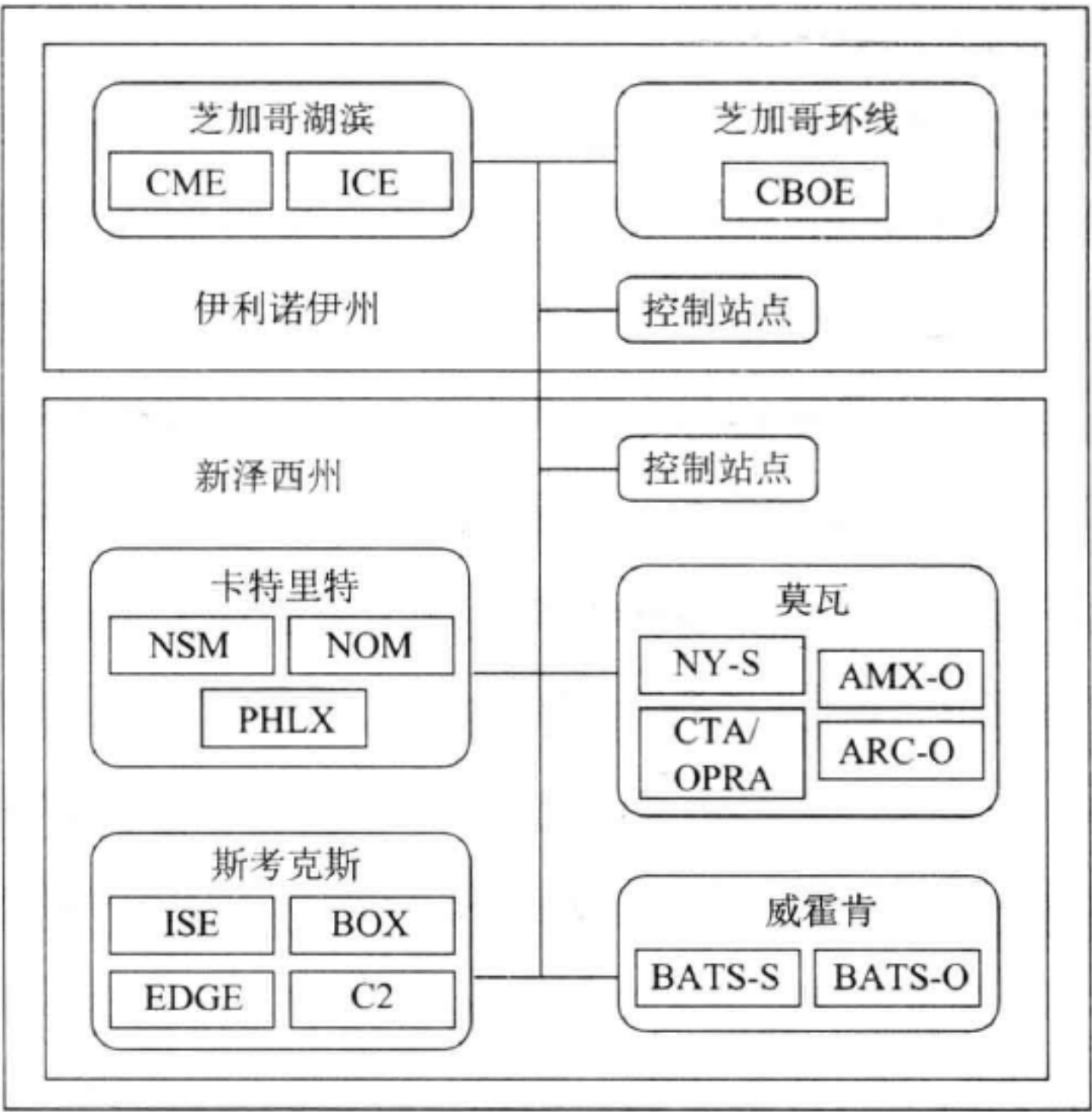


图 5-1 服务器站点示意图

基础设施：系统通信和数据库

正如我们即将看到的，这个系统的工作将最终由一系列低耦合度的计算机程序组件来完成。为了让这些组件之间能够互相沟通，我们会在系统中的每一个远程和控制服务器中植入一个 Tervela 公司的 TMX 信息交换机以及它的支持产品。TMX 交换机将作为所有需要相互通信的组件之间的信息中转站。[⊖]

要安装 TMX 服务（这个原则也适用于任何类型的基于“发布与订阅”机制的系统通信解决方案），我们需要定义一系列的“话题”，

⊖ 我们之前提到过，TMX 信息交换机是 Tervela 公司提供的一种基于硬件的“发布与订阅”机制的通信解决方案。如果要使用基于软件的通信解决方案，我们可能会选择同样具有竞争力的 29West 公司提供的 LBM 及其附属产品。

也就是消息类型，组件使用这些消息类型在彼此间进行通信。例如，我们可以定义一种叫做“交易”的消息类型，并且只要发生了一笔交易，就可以在系统中统一使用这种消息来通知其他的组件。一条“交易”的消息会包含一系列严格定义的细节，例如交易的产品种类、交易所、交易价格等。任何与交易相关的组件都会订阅这个“话题”。一个典型的高频交易系统的“话题”列表可能会很长，其中“话题”的种类可以达到数千种之多，例如理论价格的变动、新市场数据的接收、交易员调整系统参数等。

与大多数其他的计算机系统一样，一个高频交易系统也会产生各种各样的数据，并且需要保存下来：可交易证券的列表（例如股票代码表）、交易记录、持仓头寸、收盘价、交易所编号、用户密码、交易员的名字，等等。这些数据因为各种各样的原因都需要被永久保存。例如当系统重启时（可能是管理员有意为之，也可能是因为发生了某种系统崩溃），它必须知道在重启前的各种系统状态：当前的持仓头寸、人为设定的参数等；当然，各种数据挖掘和系统报告也都需要这些数据。为此，我们的系统将为每一个服务器站点上配置一个关系型数据库，例如 Sybase 和 Oracle 提供的数据库。

各种组件

现在，我们可以开始介绍系统的各种组件了（也就是各种软件程序），这也是实际执行高频交易操作的部分。如果说网络、通信和数据库这些基础设施提供了舞台和灯光，那么这些组件就是高频交易这场大戏的演员。在我们的系统中，这些组件是用 C++ 语言编写的，并且运行在 Linux 操作系统的一个定制版本上，这个操作系统已经被

优化过了，移除了所有我们不需要的部件，^① 并且被安装在多核处理器上。这些高独立性的低耦合组件将会使用我们之前提到过的 TMX 服务进行相互通信，并使用数据库在必要时对数据进行存取。

我们会做一些事情来进一步提高组件的性能。首先，我们使用 TOE 卡^②来与网络交换数据，把处理器从这种繁重而重复的工作中解脱出来，为高频交易组件的运行留出更多的时间。当然，我们还希望发挥多核服务器在并行处理能力方面的优势，把各组件的任务分配到尽可能多的处理器上去，这要求组件的程序使用多线程的方式编写。对于程序员来说，这是一项相对来说很繁重的工作，至少相对于编写单线程的程序要困难得多。幸运的是，英特尔公司提供了一个叫做“进程创建模块”(threading building block, TBB)的 C++ 运行库，它能够帮助我们处理这些麻烦的细节，在它的帮助下，我们可以节省非常多的系统开发时间(当然，还有许多令人头疼的问题)。

典型的高频交易系统的组件结构可以非常庞大。一个大型系统经常会包含数千个同时运行的组件程序实例，它们之间协同工作以完成高频交易的各种任务，这样的系统通常都具有很高的复杂性。不过，对于那些最复杂的系统而言，由于许多组件实际上都在做着非常相似的工作(如果不是完全相同的话)，因此要把这些组件分门别类并不是一件太困难的事情。^③ 比如，可以把系统中的组件分成以下 5 个大类：策略组件、监听组件、定价组件、交易组件和管理组件。图 5-2 显示了这些类别的组件如何进行交互。

策略组件的任务是将人们的思路转化为其他组件所要执行的指

① 这看起来有点像一个旅行商店为了加快速度而卸掉了一辆装着货物的卡车。

② 我们在第 4 章中曾提到过它，其全名是 TCP/IP 减负引擎卡。

③ 在一个纯粹的面向对象系统设计中，这些包含多种组件的类别通常是一些高级的基类，从这些基类中可以衍生出许多子类，这些子类的实例也就是实际运行在系统中的组件程序。

令。举例来说，我们都知道，不同的高频交易公司会采用不同的交易策略，可能某家公司只交易股票，并且会采取刮取返还（rebate scratching）策略和吃掉动作慢的（slow-mover takeout）策略。刮取返还策略是一种基于交易所交易费返还的策略；而吃掉动作慢的策略则是在市场突然大幅变动时，反应较快的做市商吃掉反应较慢的做市商还来不及更新的价位上的报单的策略）。另一家同样只交易股票的公司可能使用的是推动大象（move the elephant）和趋势跟踪（follow trends）策略；还有一家公司可能同时交易股票、期权和期货，并会尝试像波动率套利这样的富有挑战性的策略。因为我们的系统实现了股票、期权和期货的高频交易功能，它可以实现以上讨论过的所有这些策略。而策略组件除了将这些思路转化为其他组件可执行的指令之外，还能实现一些其他的简单功能，比如维护可交易证券的列表，这同样取决于人们想要交易什么。市场上有几十万种可以交易的证券——大约 5 000 只股票和 30 万种期权合约，这件看起来很小的“家务事”实际上也不是那么容易能够完成的。

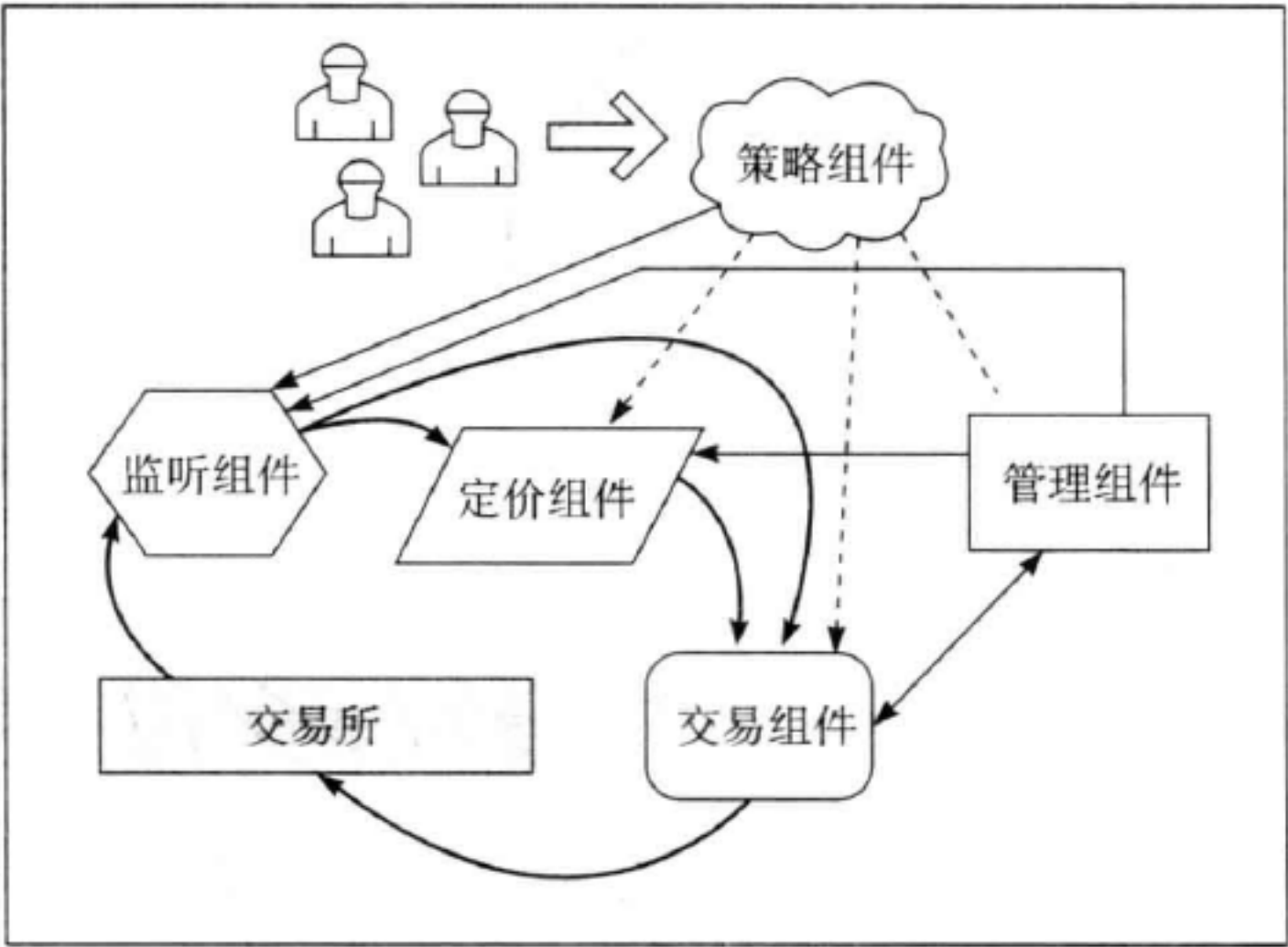


图 5-2 组件交互示意图

监听组件的任务是获取市场数据并将它们转换成其他组件能够

使用的格式。它们处在离市场数据源最近的地方，例如交易所或者CTA/OPRA平台，监听每一个市场数据，检查其中是否有错，在不同的数据源提供的数据不一致时进行仲裁，然后将经过处理的市场数据发布到系统的其他部分。当然，所有的这些任务都是在非常短的时间内完成的。

定价组件的功能应该和你想的差不多。它们会实时计算各种证券的价格，交易组件会根据这些价格向交易所下单。需要说明的是，定价组件的定价过程会受到交易潜在盈利 α 和系统的持仓水平的影响：策略组件会向定价组件发送指令，让它去寻找潜在的盈利机会；而管理组件则会根据系统的持仓水平向定价组件发送指令，调整后者的定价逻辑。

和监听组件一样，交易组件也直接与交易所通信，但是它的任务是向交易所发送委托单和报价。它们从监听组件获取市场数据，从策略组件获取交易指令，并且从定价组件获取证券的理论价格，最后，它们还会向管理组件报告交易结果。

管理组件的基本任务是根据交易行为和其他有意义的事件来控制其他组件的工作——主要是交易组件的工作。例如，当一笔交易发生时，管理组件会更新交易证券的仓位，而这个仓位会被反馈给定价组件来调整其定价逻辑。管理组件还要负责其他一些重要的任务，例如在发生了一些高风险事件或者糟糕的交易时，通知交易组件修改或者停止交易。

可以看到，我们的监听组件、定价组件、交易组件和交易所之间是通过有方向的、有些循环的相互作用连接在一起的。这就是对任何高频交易系统都非常重要的低延迟链路圈。一笔市场数据从交易所传来，监听组件会获得它，在一系列检查和处理之后再将它传到

系统的其他部分，定价组件使用它对证券重新定价，而交易组件则可能使用这个新的价格向市场报单。这些事情的处理过程进行得越快，我们的系统就越有竞争力。因此，当实现一个高频交易系统时，我们会发挥我们所能想到的任何技术或者逻辑上的优化手段的优势，来将这个过程的耗时减少几毫秒甚至是几微秒。而其他一些组件之间的通信，比如策略组件和定价组件之间的通信，因为对延迟并不敏感，所以我们也就不必花太多工夫让这种数据通路变得更快。

我们接下来将要逐一讨论上面这个组件清单中的各个关键成员。除非特别说明，否则这些组件的实例都会运行在我们所有的远程服务器上。下面的讨论并没有穷尽系统中的所有组件，一个真实的系统通常还会包含具有各种其他功能的组件，但是我们已经把所有基础的组件都包括在内了。我们对组件功能的描述只关注其最基础的工作职能，并忽略了一些特别伤脑筋的细节问题。当总结了每一个组件的具体任务之后，我们会描述一个具体的场景，来展示这些组件是如何进行交互协作，并最终完成高频交易这项工作的。

策略组件

策略组件的任务是将专业人员下达的命令转化为具体的各个其他功能组件所能够执行的指令，主要包括三种组件：策略服务器、证券列表管理器和合规管理器。

策略服务器

你将会在接下来看到，系统中大多数的组件都是高度专业化的，一般来说它们只做一件事情。策略服务器的基本功能是从专业人员（包括交易员、量化分析师、系统支持工程师）处获取命令，并将它们转化成具体的指令，下达给其他的各个组件，而这种转化过程则主要取决于下达命令的人所希望实现的策略。例如，如果公司希望

实现一个基于股价相对于股指期货价格变动的滞后效应的交易策略，那么它会让策略服务器通知股票定价组件计算期货指数中的每一只成分股由于期货价格变动而产生的 α 。而在其他组件启动的时候，策略服务器也会告诉它们之后的行为模式。从这个角度来说，策略服务器看起来很像一家依靠交易员人工交易的公司里每天的例行晨会，在晨会上，风控经理和交易员会在开盘前聚在一起，确定这一天的交易策略和交易额度限制。

在策略服务器组件的设计中，一个重要的考虑是，将策略的转化过程集中在一处，而不是将它分散到多个地方，因为这样会使得系统相对来说更容易维护。假设这些策略（比如说趋势跟踪策略）指令零碎地分散在各个地方，那么当你想要修改策略的实现时，需要在许多地方都进行修改，这很显然会导致你更容易错过一些东西。策略服务器组件的指令和其他组件之间的交互对时间并不敏感，所以只需要把它放在我们的控制服务器上就行了。

证券列表管理器

证券列表管理器组件的功能看起来并不太复杂，不过仍然是非常重要的。它记录着系统可能交易的所有证券。具体说来，它维护着一张包括股票、期货合约和期权合约的列表，这张列表还包含这些可交易证券的一些其他细节信息：对股票来说，这些细节包括预期的分红时间和收益公告时间等；对期权来说，这些细节包括到期日、执行价格之类的信息。期货合约和股票的列表可以由人工输入来完成，但是由于期权合约的数目太多了，所以最好还是使用一种更加自动化的方法来处理期权的列表。

几乎每天都会有新的期权合约上市。举个例子，假设 Acme Explosives 公司的股票目前的价格大约在 50 美元左右，那么看涨期

权的最高执行价格会是 80 美元，如果股价涨到了 60 美元，交易所一般来说就会在市场上挂出一个新的执行价格为 90 美元的看涨期权。大多数新的合约都是在隔夜的非交易时间段挂牌交易的，因此证券列表管理器必须在每天交易结束后，或者在第二天开盘前从期权清算公司（Options Clearing Corporation, OCC）^①或各个独立的期权交易所获取这些新的合约，它同样也会在日内查询这些信息源，看看是否有新的执行价格的期权挂牌。我们只需要在控制服务器上部署这种证券列表管理器组件就够了。

合规管理器

对任何交易公司来说，“合规”这两个字看起来都挺好笑的，这是指用一些法规条文来规范交易的行为。其中的多数规则主要涉及的是填写报告、交易员的交易资格，以及诸如此类的与高频交易系统无关的事情。不过还是有一些规则会对高频交易过程中的实时行为产生影响，这些规则必须在系统中体现出来。

一个具体的例子是，一些期权交易所对期权做市商有“合法价差”的要求。这种要求对做市商在一个期权上的买卖价差进行限制。大多数期权交易所在正常的交易日都会要求期权做市商所报的买卖价差不得超过 5 美元，不过这个价差实在太大了，所以大多数时候它都没什么实际意义。合法价差的范围则窄得多（至少在交易日的开始时段内），另外，它还会根据期权的到期时间和申买报价进行调整。合规管理器维护着所有诸如此类的规则细节，并且在必要时向对应的组件提供这些规则（在合法价差规则的例子中，合规管理器组件会向系统的做市组件提供这些信息）。同样，合规管理器组件只需要

① 期权清算公司目前由各期权交易所共同拥有，其职责是为期权交易提供清算和结算服务。

部署在系统的控制服务器上。

监听组件

我们的系统中有四种监听组件，其任务是从外部收集市场数据和信息，并且将它发送给系统中的其他组件，它们分别监听股票市场数据、期权市场数据、期货市场数据和各种新闻。

股票市场监听器

股票行情监听器组件通过原生 API 与交易所连接，并监听股票市场的行情——当前的申买价、申卖价以及各种其他交易信息，这些行情既由交易所独立发布，同时也会被统一数据平台转发，然后将可疑的数据过滤掉，把经过过滤的数据以广播的方式发送给系统的其他组件。对可疑数据的过滤是非常关键的。每天市场上会有数亿条数据被发布，其中的一些有可能因为人为的或技术上的各种原因出错。你决不会希望你的系统被这些有瑕疵的数据所误导，向市场进行错误的报价或下单，这会让你亏掉很多钱。需要注意的是，我们发布的不仅仅是委托单簿顶部的行情，而是全市场深度行情。股票监听器还发布所有交易该股票的市场的状态（期权和期货监听器也是如此）。

任何一个交易市场都有好几种可能的当前状态。例如，在晚上，大多数的交易所处于收盘状态，而在白天的很长一段时间内，它们处于开盘状态。显然，你不希望你的系统在交易所收盘时段内去做任何交易。另一种常见的状态是速动市场（fast market），这种状态通常出现在市场经历了突然的非正常大幅波动的时刻。在此状态下，交易所可能会对一些交易规则做出修改。监听组件在这时需要向系统发送市场状态的变化，让其他组件能够对这种变化做出反应。

期权市场监听器

期权市场监听器组件的工作和股票市场监听器并没有多大不同，只不过它的数据源是各个期权交易所和 OPRA 平台。因为期权市场上的合约数目远远超过股票市场，因此期权市场监听器会比那些监听股票市场的同伴们忙得多。

期货市场监听器

与股票以及期权市场监听器一样，期货市场监听器组件监听、清洗并向系统中的其他组件发布各种股指期货行情，比如我们之前提到过的 E-mini 标普 500 指数期货合约（ES）。除此以外，这个组件同样会发布来自于芝加哥商品交易所的欧洲美元期货合约行情数据。欧洲美元期货合约的持有者可以在未来某段时间以约定利率借入或者拆出一笔金额。*LIBOR* 是伦敦银行间拆借利率（London interbank offer rate）的缩写，它在市场中广泛地用于无风险利率的基准，也是期权定价的输入参数之一。我们接下来会看到，一个叫做“利率曲线生成器”的组件会使用这些欧洲美元期货合约的价格来推导出无风险利率。

新闻监听器

市场新闻监听器组件的任务有点像一个坐在电脑前盯着彭博终端屏幕的交易员，它会不停地搜索一些能够引发市场兴趣的关键词，比如一家上市公司发布公告改变了之前的股利分配方案，或者发布了一份新的盈利公告，这样的信息会对股票价格或者期权价格产生影响。一旦探测到了这种信息，市场新闻监听器就会马上把它广播给系统中其他可能会对这些信息做出反应的组件。例如，一个用来实现事件驱动交易策略的组件很显然会订阅这样的新闻，还有我们之后将要提到的期权定价组件，同样也需要订阅这些信息来获

取关于股利变化之类的新闻，从而修改股利预期并更新对期权的定价。

定价组件

我们接下来讨论的这些组件会直接或间接地为我们所要交易的证券进行定价。

隐含波动率生成器

股票或股指的波动率是对于其价格变化的频率和幅度的一种数值上的描述，它同时也是期权理论价格的计算中最重要的因素之一。对于波动率这个概念，最困难的地方是，虽然有很多地方需要用到它，但没有直接的方法可以测量它。[⊖] 不过你可以从当前的期权价格反推出一个隐含的股价波动率，而这就是我们所讨论的这个组件一直在做的事情。隐含波动率生成器会一直监控市场上期权的当前价格——申买价、申卖价以及上一笔成交价，并且用一个期权定价公式（例如布莱克－斯科尔斯公式）来反推出股价的隐含波动率。

也许这听起来有点奇怪，不过我们会做解释。对一个期权定价公式而言，标准的输入参数包括当前的股价、预期股利、执行价格、距到期日的时间、无风险利率，当然，还有波动率。如果要根据公式计算出期权的理论价格，那么你要做的就是搜集所有的这些参数，然后把它们放到公式中去。但是如果你已经看到了市场上的期权价格，同时还知道除了波动率之外的其他几个参数，那么你就可以把公式改写一下，然后反推出要得到这个价格的波动率应该是多少，

⊖ 当然，你可以很容易地计算出一种资产的历史波动率，但是对期权定价需要的是即时的波动率，而不是之前一个月或之前一秒钟的波动率。

这就是隐含波动率的概念。[⊖]

波动率曲线生成器

上面已经提到，标准的期权定价公式中需要波动率这个参数。对于一个期权交易者，特别是期权做市商而言，无论怎么强调波动率的重要性都不为过。在期权交易中，可以说波动率就是期权的“价格”，而那些几块几分的权利金之类的东西是第二位的。期权交易者的脑子里经常想的不是买卖期权，而是买卖波动率。如果一个高频交易系统以 1 美元的价格买入了 100 股某公司的股票，并很快地以 1.02 美元的价格把它们卖了出去，那么看起来这是一笔不错的股票交易；而对于高频期权交易而言，一笔好的交易是，系统以 30 点的波动率买入 100 份期权，然后很快以 32 点的波动率将它们卖出。

波动率曲线生成器的工作是持续地计算并向系统发布由不同执行价格和到期时间的期权价格根据期权定价公式所反推出的隐含波动率，主要是为期权定价组件服务。在这里，这些期权的隐含波动率以曲线的形式表现出来。最常见的一种曲线是一条简单的正交坐标轴曲线，x 轴（水平轴）表示期权的执行价格，而 y 轴（纵轴）则表示波动率。图 5-3 中画出了一条这样的曲线。

这条曲线显示，任意一个执行价格上期权的隐含波动率受到两个输入参数的影响。一个是隐含波动率，还有一个是我们手头的存货，即在哪些执行价格上的期权的仓位。回想一下最基本的倾斜做市策略。该策略的基本思路是要完成双向交易（也就是一买一卖，或者一卖一买），为此，在一笔交易之后，需要修改你的申买价（如果刚

⊖ 芝加哥期权交易所的网站上提供了通过一个理论价格和其他几个公式参数来计算隐含波动率的小功能（cboe.com/LearnCenter/OptionCalculator.aspx），当然，它的目的只是进行投资者教育。

刚卖出)或申卖价(如果刚刚买入),让你的报价在市场上更有吸引力,从而增加反向交易完成的概率。

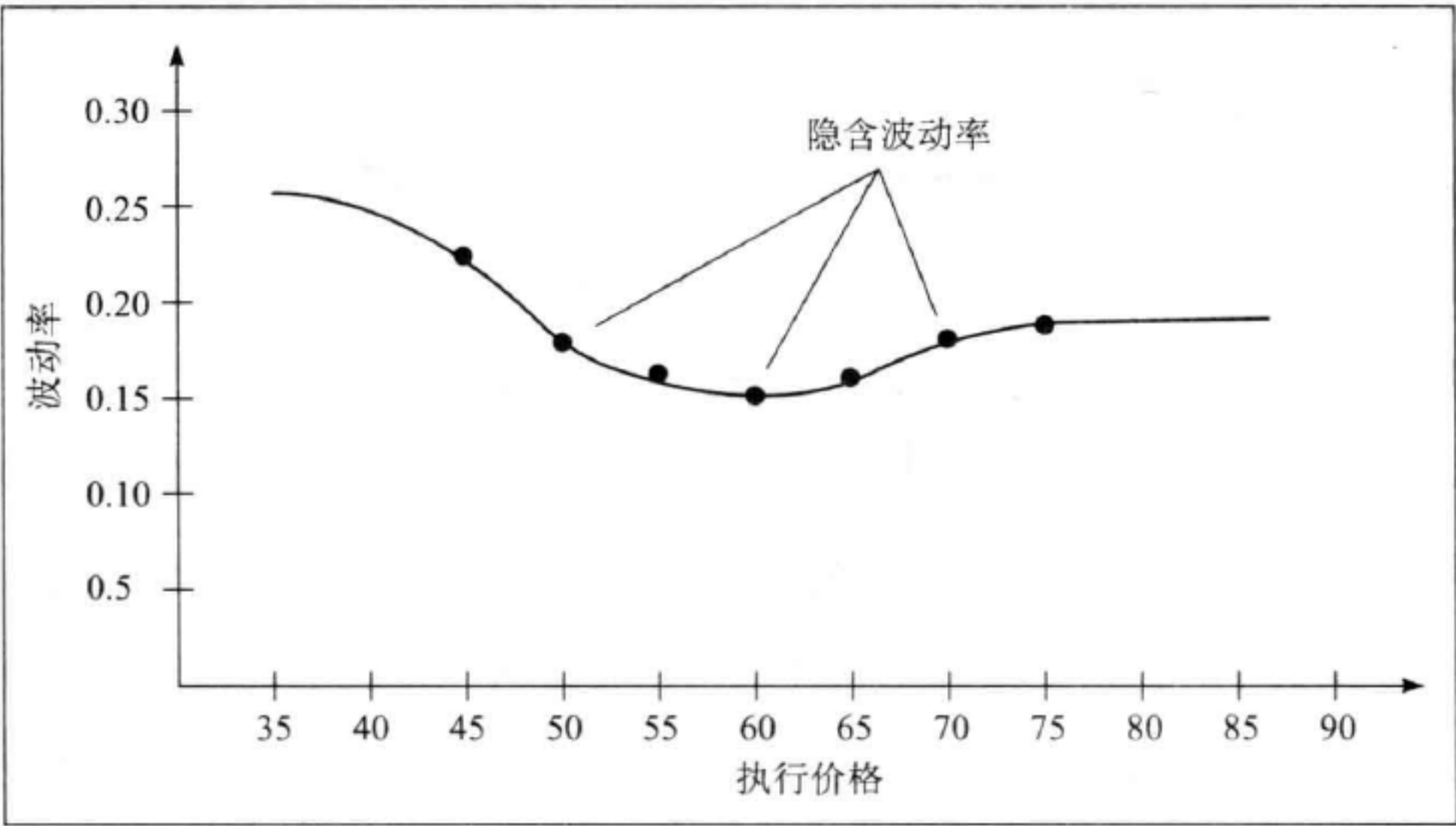


图 5-3 波动率曲线

在期权交易中，这种倾斜策略不仅可以通过修改实际的期权报价来完成，还能够通过修改波动率曲线的方式来完成。对于期权来说，这种修改波动率曲线的方式尤其方便，因为实际上会有好几个期权（例如同一个执行价格上的看涨期权和看跌期权）在计算理论价格时使用相同的波动率，因此修改一个波动率实际上同时修改了多个期权的报价。如我们刚才所说的，波动率就是价格。如果你已经买入了一些某些执行价格上的期权头寸（买入波动率），为了完成反向的交易，你完全可以通过调低在相应执行价格区域的波动率来让你成为一个有竞争力的卖家。

如果你已经非常习惯地把波动率看作期权价格的一种形式，也许很快就能想出第三个影响波动率曲线的因素，它非常类似于股票中 alpha 的概念。考虑两条波动率曲线，其中一条是基于标普 500 指数的期权波动率曲线，而另一条则是基于纳斯达克 100 指数的期权波动

率曲线。假如你相信（一些高频交易者确实是这么认为的）这两条曲线上的波动率存在领先-滞后效应（其中一条曲线上的波动率水平发生改变，引发了另一条曲线上的波动率随后发生同向变化）。在某个时刻，基于标普 500 指数的期权的隐含波动率曲线因为某种原因突然上升，那么你应该非常迅速地向上调整另一条基于纳斯达克 100 指数的波动率曲线，来抓住这个非常短暂的相对低估的机会买入波动率，然后当市场做出向上的调整之后，再卖出波动率以实现获利。

图 5-3 中的曲线给出了一些给定的不同执行价格上的期权隐含波动率，而另一种常见的波动率曲线刻画的则是不同到期时间的期权隐含波动率。波动率曲线生成器既可以将这两个维度上的波动率曲线分别发布，也可以把它们合在一起形成一张波动率曲面。图 5-4 给出了一张波动率曲面的形状，我们将在后面用它来讨论波动率套利策略。

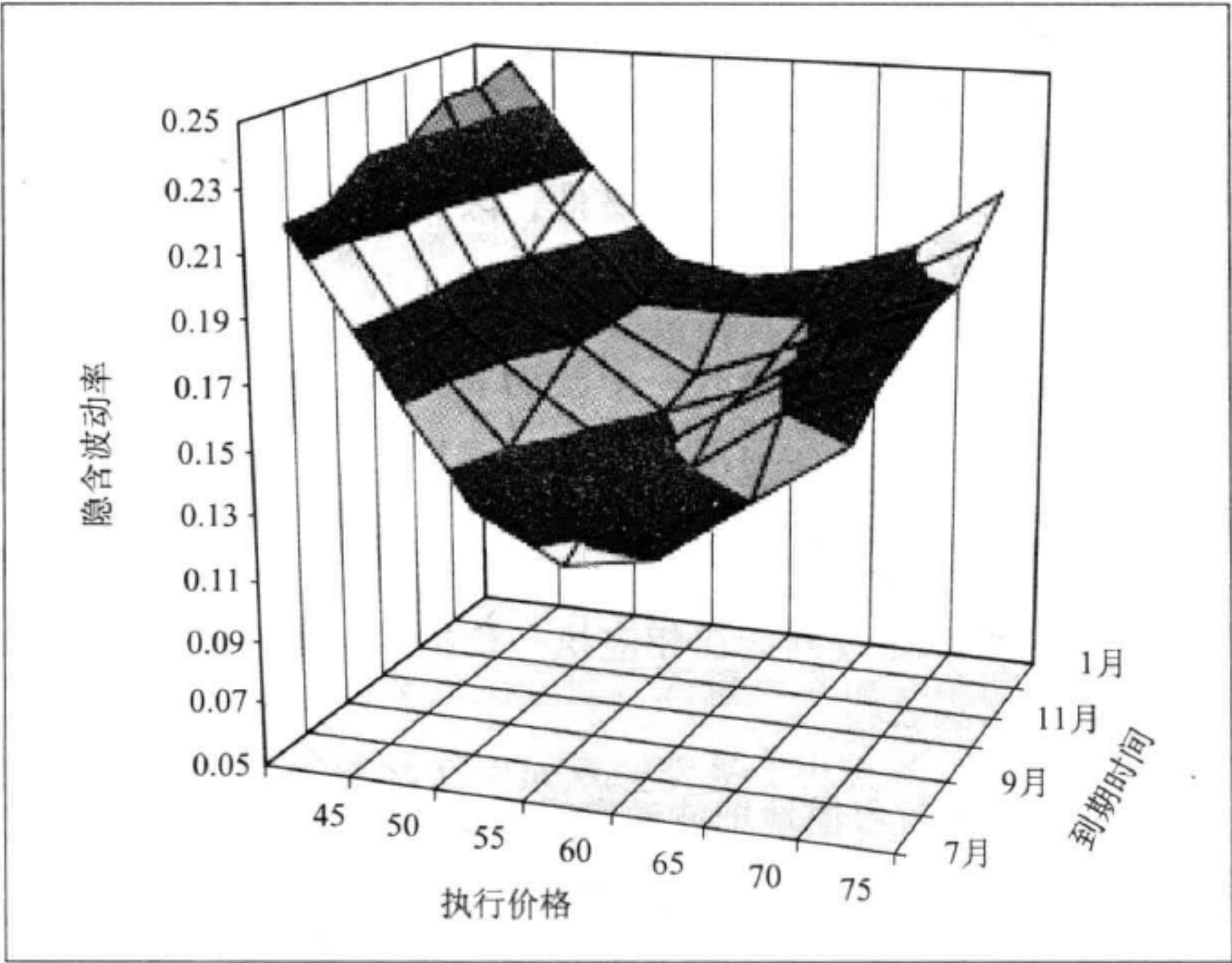


图 5-4 波动率曲面

利率曲线生成器

任何一个标准的期权定价公式都需要无风险利率这个输入参数，利率曲线生成器组件所做的就是持续地计算并且向系统中发布利率曲线（也叫做收益率曲线）。和波动率曲线一样，利率曲线也画在一个很简单的正交坐标轴中，显示不同期限上的利率水平。图 5-5 给出了一条利率曲线的示意图。

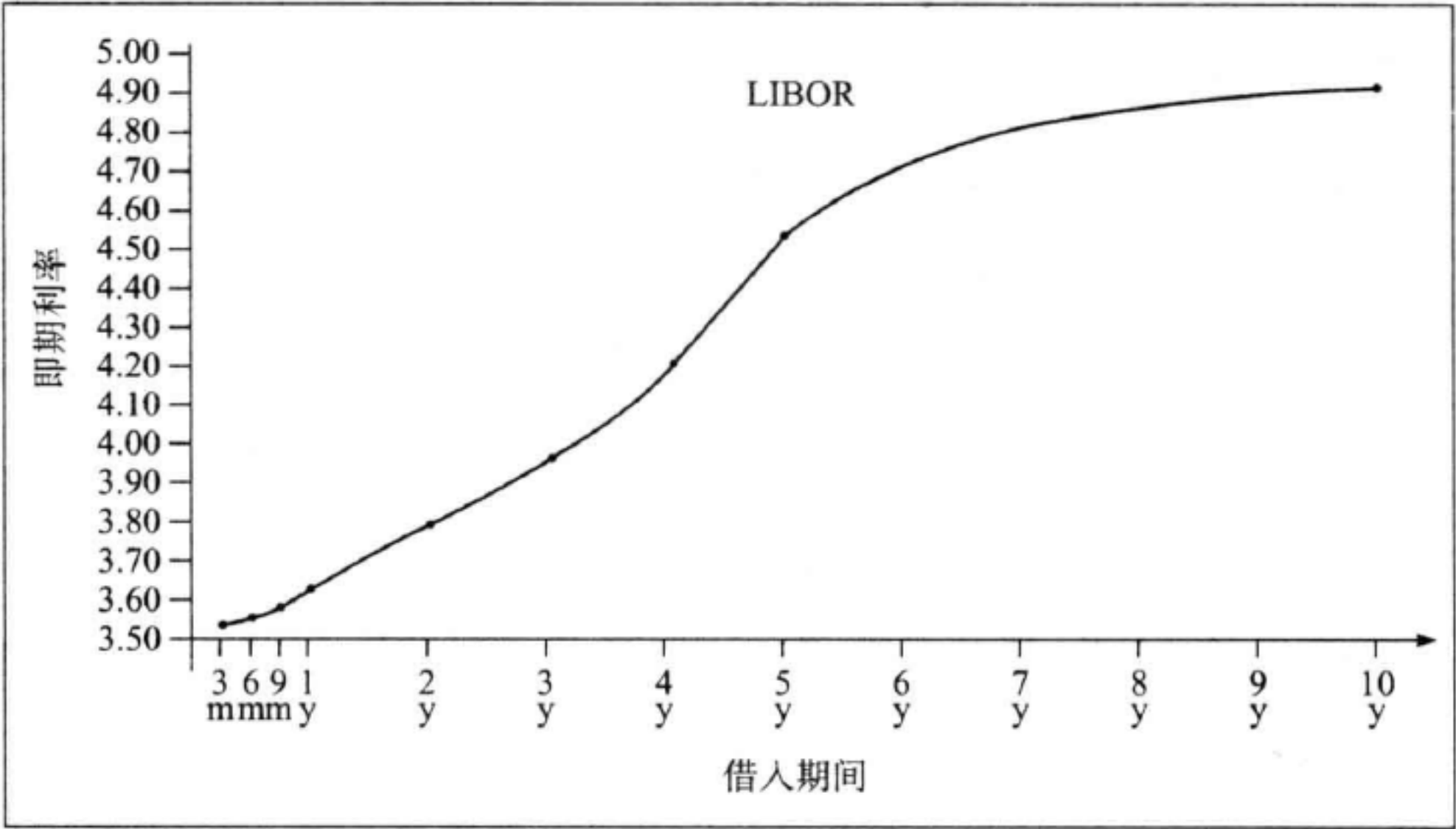


图 5-5 利率曲线

利率曲线生成器主要使用当前欧洲美元期货合约的市场价格来计算利率水平，我们在前面提过，欧洲美元期货合约在芝加哥商品交易所进行交易，期货市场监听器会将它的行情发布到系统中[⊖]。这种计算利率的过程有点乏味：一张欧洲美元期货合约锁定了以未来某个时点为开端的一段时间内的借款水平，因此这种期货合约的价格中也就隐含了远期利率，我们可以从中得到一个相对应的即期利率，而这也就是我们所想要的。利率曲线生成器组件会在每收到一个欧

⊖ 除了使用欧洲美元期货合约的价格之外，我们也可以监听美国国库券的价格信息，并使用它来计算利率曲线。

洲美元期货报价时重新计算即期利率，然后将它插入到固定到期时间的利率当中，形成类似图 5-5 中的一条利率曲线。

股票定价组件

股票定价组件是我们的系统中一个相当重要的组件，它计算所有经过 α 调整的股票价格。计算这些价格的基础是股票的当前市场价，股票市场监听器会将它们发布到系统中，而计算 α 的方法则有很多（我们在之前章节中讨论配对交易策略时曾经提到过这些方法）。举个例子，在一个基于股价对股指期货变动滞后效应的策略，股票市场监听器会监听由期货市场监听器发送的标普指数期货合约的价格。根据指数期货合约价格变动的方向和幅度，以及之前观察到的股价变动与指数期货价格变动之间的关系（例如 β ），定价组件计算出一个 α 因子，并使用这个因子对当前的股票定价做出修改。

期权定价组件

对于每一个我们想要交易的期权合约，期权定价组件都会为它计算一个当前的理论价格，以供做市组件使用。这个理论价格基于当前的股票价格（由股票定价组件提供）、波动率（由波动率曲线生成器提供）、无风险利率（由利率曲线生成器提供）、预期股利、执行价格和距离到期日时间（最后三个参数由证券列表管理器提供）计算得到。这个组件可以使用很多种期权定价公式来计算期权的理论价格，当然，其中最著名的一种还是布莱克-斯科尔斯期权定价公式，但是它只能用于计算欧式期权的理论价格（欧式期权只能在到期日执行）。对于许多美式期权而言（美式期权可以在到期日之前的任何一天执行），它们的定价需要使用一种二叉树模型，还有另一个基于树结构的模型——三叉树模型，它有一个很相似的“近亲”叫

做有限差分法（finite difference method, FDM）。对于这些模型的进一步探讨超出了本书的范围（市场上有很多书对这些模型的讲解都很不错）。在这个假想的系统中，我们将使用有限差分法为美式期权定价。它是一个非常高效的模型，同时还能够与计算机的并行处理配合得很好，而并行计算也是我们将尽可能多使用的技术之一。

不论使用哪一种公式，在股价变动后对期权价格的重新计算都需要花上一些时间。为了在交易中减少这种时间开销，我们将使用一种在很早以前被用在人工期权交易中的技巧。由于变动最频繁的期权定价参数是标的股票的价格和其波动率，我们会在每个交易日开始之前以标的股票当前的价格和波动率水平为中心，预先计算期权在一系列可能的价格和波动率之下的理论价格，对于所有不同执行价格和到期时间的期权，我们都会重复这项工作。那么到了标的股票价格真正发生变化时，期权定价组件就不需要重新计算其理论价格了，而只需要查询一个含有所有这些预先计算过的理论价格的三维矩阵就够了。

这种方法在之前就被人工期权交易员们广泛使用。每天早上，这些交易员和他们的书记员都会打印出一厚摞这样的单子，这些单子上记录着标的股票在不同价格和不同波动率之下期权的理论价格，这些交易员会在交易时段不停地查这些单子（当然，如果股票的价格变动超出了预先计算的范围，他们也会收到一些新的单子），这样他们就可以方便地获得期权的理论价格，而不需要花太多时间重新计算。计算这样的一些矩阵实际上并不像看起来那么容易。首先，即使是预先计算也要花掉不少时间，你不可能为所有可能的标的股票价格和波动率水平都进行预先的计算，事实上你只是选择一些比较合理的、离散的价格和波动率水平，并计算对应的期权理论价格，

然后通过插值的方法得到在所有的可能的价格和波动率水平下的期权理论价格。另一个挑战在于，插值不能太少，否则它会降低插值所获得的价格的准确率；如果插值的频率太高，又会使得计算所花的时间太长，这也违反了你的初衷。图 5-6 给出了一个预先计算出的价格矩阵[⊖]。

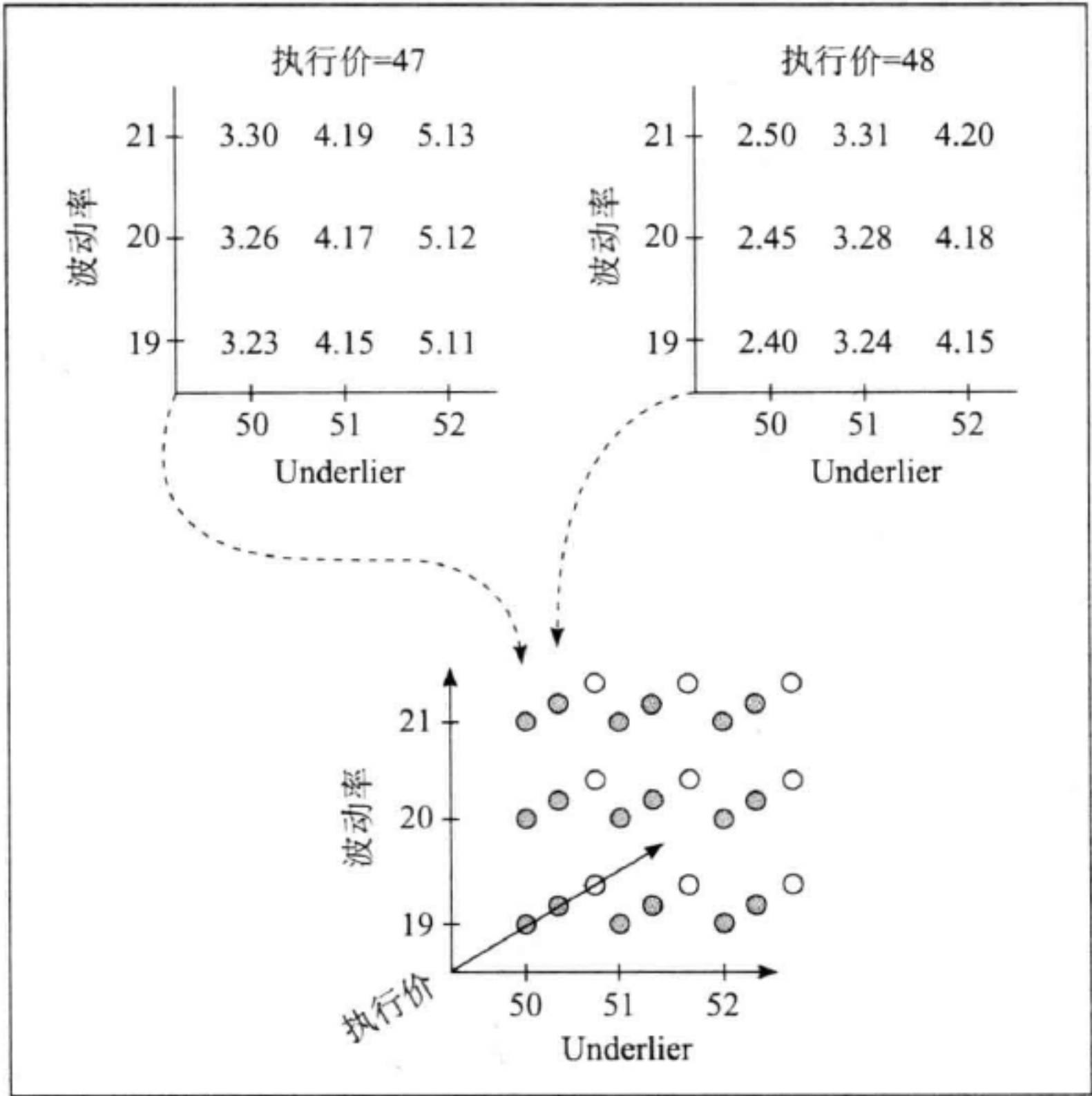


图 5-6 三维期权价格图

做市组件

做市组件的功能是为系统进行做市的每一种股票和期权合约计算出一个买卖价差。计算买卖价差的依据包括：当前的市场状态、策

⊖ 期权价格是使用芝加哥期权交易所提供的网上期权理论价格计算功能得到的 (cboe.com/LearnCenter/OptionCalculator.aspx)。这是一个美式期权，标的是一只无股利的股票，距到期日时间为 30 天，年化无风险利率为 2%。

略服务器的指令，以及股票定价组件和期权定价组件的定价水平。股票的做市组件同时还会考虑当前的持仓水平（存货），并且据此对市场做出倾斜调整。我们之前提过，期权的倾斜调整已经包含在生成波动率曲线的过程中，因此做市组件中不再进行这样的调整。此外，期权的买卖价差还会考虑一些其他因素，比如对冲交易的成本、融资成本（例如，借款买入期权时的成本），以及所谓的“波动率的波动率”（回想一下，波动率实际上是期权价格的一个代理变量，在一个波动率变动频繁的期权和波动率比较稳定的期权之间，做市商会为持有波动率变动频繁的期权要求更多的风险补偿）。

交易组件

系统中的交易组件包括四类：报价引擎、电子眼（包括一种专门执行期权价差交易的组件）、股票定位器和对冲器。它们直接与交易所进行交互，完成交易执行工作。

股票报价引擎

股票报价引擎的功能是将做市商组件对股票的报价（包括买价、卖价和相应的股票数量）传递给交易所。和监听组件一样，报价引擎组件也是使用交易所的原生 API 编程实现的，因此，我们在每一家交易所的报价引擎组件都是不同的（例如，不使用 FIX 协议）。虽然向交易所的报价引擎传送报价这个概念看起来很简单，但是在实践中它却是一项非常冗长复杂的工作。各交易所在很多方面都存在着一些细微的差别，例如在一个报价包中最多能容纳多少个报价、接收报价的频率、如何向报价引擎发送收到报价的信号等。把这些细微的差别都弄清楚是非常重要的，因为对做市商而言，报价是和交易捆绑在一起的义务，并且报价几乎时时刻刻都需要更新。如果你想改变报价，但却因为某种原因失败了，你很有可能会眼睁睁地

看着别人把你来不及更新的报价给吃掉，强迫你做出一笔坏的交易。

股票报价引擎的另一项任务是对安全控制进行正确设置，大多数（但不是全部）交易所都有这样的控制策略，以保证做市商不会做出一些他们并不想要的交易。例如，交易所可能会让你设定 X 和 Y 的值，如果你在 Y 秒内完成了 X 笔交易，交易所就会自动地将你在市场中的所有报价都置为无效。另一个常见的例子则被称为“心跳监视器”，这里的概念是，报价引擎每隔一段时间向交易所发送一条简单的信息（被称为一个“心跳”）来告诉交易所这个引擎是能够使用的，或者说“活着”。如果交易所探测不到你的“心跳”，那么它将认为你那边出了问题，并将你的报价从市场中移出。

股票电子眼

股票电子眼（electronic eye，EE）是一个与股票报价引擎很类似的组件，它根据市场的买卖价差和数量以及策略服务器的指令向某家特定的交易所（使用原生 API）发送委托单（举例来说，策略服务器可能会向电子眼发送指令，告诉电子眼是否使用比报价引擎更紧的报价）。不过报价引擎向市场发送限价单（也就是它们的报价），电子眼则发送不成即撤委托单。报价引擎只发送被动报单，电子眼则进行主动交易。

期权报价引擎

与股票报价引擎相同，只不过是期权进行报价。

期权电子眼

与股票电子眼相同，只不过用于期权交易所。

期权价差策略执行器

与期权电子眼一样，期权组合策略执行器下单的类型也是“不成即撤单”，但是它不只对一个期权合约进行下单，而是对一组标准

的期权合约组合进行下单，这类组合称为价差或者策略。一个单一的期权合约通常被用来实现一些简单的投资目标，例如对股票价格可能的剧烈下跌进行保险等；而价差组合则可以用在诸如对股票价格可能的超出某个价格范围的变动——涨过某个高价或者跌穿某个低价——进行保险之类的目的上，换句话说，也就是对潜在的高波动率进行保险。一个典型的例子是跨式策略（我们之前在第3章中曾讨论过），它由相同执行价格上的一个看涨期权和一个看跌期权组成，这种组合策略能够用来降低整个投资组合的波动率风险暴露（一个叫做 vega 的风险度量）。期权价差策略执行器会监控价差委托单簿，例如芝加哥期权交易所的复杂委托单簿（complex order book, COB），并且根据策略服务器的指令进行交易。

股票定位器

要卖空一只股票，交易者可以从别人的账户上借到股票，并且在某个时刻将这些股票买回来归还给借方。卖空股票有很多种可能的原因，对股票高频交易者而言，其中的一种原因是，他想要吃掉一张市场上的买单，但是账户上并没有对应的股票；对于期权高频交易者而言，他可能想要卖空一只股票来进行期权的 delta 对冲。不论你因为什么原因需要卖空股票，原则上（至少这是一种好的做法）你都需要在卖出之前先借入这些股票。[⊖] 股票定位器组件的任务就在于此，它使用与股票经纪商（借入股票的通常来源）之间的连接来保证在卖空之前有足够的股票可以借入。我们只需要在控制服务器上配置这个组件就够了。

自动对冲器

回想一下之前描述的跳过 delta 策略，我们讨论过一个期权做市

⊖ 不借入股票的卖空行为被称为裸卖空，它是一种广受争议的操作方式。

商在一笔期权交易之后如何通过交易一定数量的股票来对期权头寸进行 delta 对冲。自动对冲器组件的任务就是进行这种对冲操作，其最基本的原理就是在一笔期权交易之后，根据在当时计算出的期权的 delta 值买入或者卖空一定数量的股票。在实践中，还会有一些别的考量。首先，也是最重要的，期权的 delta 值像它的理论价值一样，同样会随着时间改变，因此，对冲的头寸必须根据 delta 值的改变不断地进行再调整，这就要求自动对冲器持续地监控期权头寸的 delta 值变动，并且以此为根据买卖股票来保持组合的 delta 中性。同样，当一个期权到期、执行或者被执行^①时，以及期权交易不成功或者被调整（参见“管理组件”一节中对仓位管理器的讨论）时，也需要对相应的股票头寸进行改动。

我们可以很清楚地知道，当期权做市商试图按照这种常规的方式进行 delta 对冲时，会给那些试图抢在他们对冲之前交易股票的高频股票交易者带来机会。因为这个原因，一些期权做市商并不使用标的股票，而是使用一种指数合约来进行 delta 对冲（例如我们之前曾提到过的 Spider ETF，如果标的股票在标普 500 指数中的话）。这么做需要一点技巧，因为使用 ETF 进行时，用于对冲的 ETF 数量不仅取决于期权的 delta，还取决于标的股票的 beta 值。^②但是，如果做市商不希望被这些跳过 delta 交易者占便宜的话，他也没有别的选择。

除了 delta 对冲之外——也就是试图抵消由期权价值相对标的股

① “被执行”（assign）这个概念仅仅是对期权空头头寸的持有者（也就是期权卖方）而言的。当一个期权多头头寸的持有者（期权买方）决定执行期权，交易所将会选择某个期权空头头寸的持有者来履行他的义务，按事先确定的执行价格向执行者卖出或买入相应资产（分别对应于看涨和看跌期权），对于这个被选择的空头头寸持有者，他的期权头寸称为“被执行”了。

② 回想一下，beta 值度量的是股票价格相对于市场的整体价格变动的幅度，在实践中，一般会使用一个包含这只股票的指数来代表市场整体价格。

票价格变动的敏感性所导致的组合价值变化——自动对冲器还负责进行 ρ 对冲。 ρ 是期权价值相对于无风险利率变化的敏感性的一个测度，当无风险利率变化时，期权价值也会发生变化。交易者可以使用一种很类似于 δ 对冲的方式来对冲这种风险暴露，而这种对冲所使用的工具则是欧洲美元期货合约（我们在介绍利率曲线生成器时提到过它），所使用的因子是 ρ 。 ρ 和 δ 一样，会随时间发生变化，因此需要重复地进行计算和对冲操作。

管理组件

以下将要讨论管理组件，这些组件并不会参与到高频交易的工作中，但是它们所负责的监控以及报告任务仍然是非常重要的。

熔断器

在任何一个交易时刻，高频交易系统在美国各个交易所中会有超过 100 万个报价，每一个报价都代表着一个以该价格进行交易的义务。当然，我们的系统会在一天中持续不断地修改这些报价来保证只在我们想要的价格上进行交易。这些报价本身就会让我们暴露在很大的风险之下，如果我们不能成功地以最快的速度来更新这些报价的话，就会有人非常愿意以一些对我们来说很差的价格和我们进行交易，这也就是“捡我们的便宜”。而我们现在拥有一个如此复杂的系统，可能会有非常多的原因导致我们在市场中报出很差的价格。市场价格的变化非常之快，实际上在被别人“捡便宜”之前，没有足够的时间来让人们通过监控发现这些问题。所以，我们需要一个专门的系统组件来监控我们所能想到的任何可能导致较差报价的状况，并且在这种状况出现时自动通知其他组件相应地调整各自的行为。

这种状况的一个例子是丢失市场数据连接。在不知道当前市场价

格的情况下，你几乎是不可能知道应该以什么样的价格进行交易的。当熔断器检测到市场数据连接丢失时，它会向所有的交易组件发送消息，并告诉它们停止交易。以自动报价组件为例，在收到这个消息后，它可能会将市场中所有的报价都撤回，也可能暂时性地将买卖价差拉大以保证这些报价不会成交。其他的一些情形包括系统触及了持仓限额、异常迅速的成交（例如在 Y 秒内交易了超过 X 张合约或股票）、发生了坏的交易以及系统的一些重要组件（例如监听组件或定价组件）功能失效等。

交易日志

这个组件的任务就和它的名字所显示的那样简单。它维护着一份系统所做的每一笔交易的日志。对于每一笔交易，它会记下交易的证券名称、交易价格、买卖方向、交易所名称以及交易的对手方（如果交易所提供的话）等各种信息。

仓位管理器

和交易日志一样，仓位管理器的任务也非常简单，但却非常重要。仓位管理器只是监听交易回报，为每种交易的证券维护一个动态仓位，并且在这些仓位发生变化时向系统进行报告。需要注意的是，并不是只有交易才会导致仓位的变化。对于交易所而言，还会不时会发生在一个交易日结束时通知某个交易者其交易被取消（这笔交易会被抹掉，就像它从未发生一样）或者调整（这意味着交易量或者价格被修改了）的事情。交易取消和调整通常是由交易所自身或者交易一方的“明显错误”导致的。[⊖] 对于期权的仓位而言，仓位

⊖ 大多数交易所都有关于“明显错误”的条款，它允许交易的一方向交易所申请取消或者调整交易。这种申请一般必须在交易发生后规定的时间之内提交到交易所。当然，如果一个职业交易者经常提出这种请求，让对方宽恕自己的错误的话，他在市场上的名声就不会太好。

管理器还需要同时考虑期权到期、执行和被执行等情况，它们都会导致期权仓位的变动。

风险管理器

高频交易者最基本的要求就是，以一个价格买入，并以一个更高的价格卖出。风险管理器所要考虑的就是在买入后卖出前所发生的事情，也就是高频交易者在等待双向交易完成时所面临的金融风险。它会不停地衡量我们持有的头寸所面临的风险。假设一个高频交易者想要以 1.00 美元的价格买入 PDQ 公司的股票并以 1.01 美元的价格卖出，现在我们看看表 5-2 所示的交易过程。

表 5-2 交易风险场景 1

时刻 1	以 1.00 美元的价格买入 500 股 PDQ 公司股票	风险头寸 500 美元
时刻 2	以 1.02 美元的价格卖出 400 股 PDQ 公司股票	风险头寸 100 美元
时刻 3	以 1.00 美元的价格买入 100 股 PDQ 公司股票	风险头寸 200 美元
时刻 4	以 1.00 美元的价格买入 400 股 PDQ 公司股票	风险头寸 600 美元

在第 1 笔交易中，交易员花了 500 美元买入这只股票。和所有的股票投资一样，这 500 美元现在面临着风险。如果股票的价格下跌，交易员会因此亏损，有可能亏掉全部 500 美元。在第 2 笔交易后，他的仓位变成了 100 股，因此他面临风险的资金也变成了 100 美元（他在这笔交易中所赚得的利润与当前面临的风险无关）。在第 3 笔交易后，他的风险暴露变成了 200 美元，在第 4 笔交易后继续增加到 600 美元。风险管理器的主要目的就是跟踪我们交易的每只股票的风险暴露。不过我们的风险管理器组件做的还不只是监控这些风险暴露。当某只股票的风险暴露超过了一个事先设定的阈值，它会自动发出交易指令来减少它。假设我们对 PDQ 公司股票的风险暴露阈值为 800 美元，让我们接下来看看表 5-3 中的交易序列。

表 5-3 交易风险场景 2

时刻 5	以 1.00 美元的价格买入 100 股 PDQ 公司股票	风险头寸 700 美元
时刻 6	以 1.00 美元的价格买入 200 股 PDQ 公司股票	风险头寸 900 美元
时刻 7	以 0.99 美元的价格卖出 100 股 PDQ 公司股票	风险头寸 800 美元

在时刻 6 的交易之后，风险管理器识别出我们在 PDQ 股票上的仓位已经超过了 800 美元的阈值水平。于是在时刻 7，它自动卖出了 100 股，这笔交易让风险暴露回到阈值，但也导致了一定的亏损。在上面的场景中，我们在这里仍然有 7 美元的净收益（在第 2 笔交易中盈利 8 美元，减去第 7 笔交易中亏损的 1 美元）。风险管理器主要关心的是管理风险暴露。如果需要以亏损 8 美元甚至更多的代价卖出股票，它同样会执行这种降低风险的交易。

对期权而言，风险管理器的任务要更加复杂一些。股票投资组合只有一个风险来源：股票市场价格的变化。而期权投资组合则不只是对股票的市场价格的变动敏感（用期权的术语来说，这种敏感性表现在 Δ 上），对其他一些因素的变化也同样敏感。期权组合的其他风险暴露由其余的一些希腊字母给出： Γ （组合价值对股价变化的二阶导数）、 vega （组合价值对波动率变化的敏感性）、 θ （组合价值对时间变化的敏感性）以及 ρ （组合价值对无风险利率变化的敏感性）。我们知道，自动对冲器承担了 Δ 对冲和 ρ 对冲的任务，尽管如此，风险管理器还是会持续计算所有的这些风险暴露，至少将它们显示给交易员和风险管理员，并将这些风险暴露值发送给系统中的其他组件（例如熔断器）。

除了根据当前的市场状态计算风险暴露之外，风险管理器同样会持续地进行一些假设分析，在各种情境下计算风险度量值。例如，它很可能会在一系列可能的标的股票价格（既可能高于市场价，也

可能低于市场价)下计算股票和期权的 δ 值。这样,交易员就能看到它们的风险会怎样根据市场的变动发生变化。这个组件很显然是系统中最不可或缺的组件之一,因此,我们不仅要将它们部署在控制服务器上,还要将它们部署在每一个远程服务器上。

盈亏计算器

毫无疑问,我们会不停地重复问一个问题:我们在赚钱吗?盈亏计算器工作就是回答这个问题。最基本的每日交易盈亏计算是这样进行的:对于每一种交易的证券,计算其在交易日结束时的价值和交易日开始时的价值之差。而对于那些当天有交易的证券,使用交易价格来替代交易日开始时的价格(因为你在交易日开始时并不持有它),然后减去所有的成本和开销(交易费用等)。如果你想在日内计算盈亏,那么就使用当前价值替代交易日结束时的价值。计算结果就是这种证券的当日盈亏,把它们合在一起就是你当天的交易操作的整体盈亏。过程就是这样,我们不需要在远程服务器上做这些事情,所以只要把这个组件部署在控制服务器上就行了。

在大多数交易公司,有两种计算盈亏的方式,它们有时被称为“前台盈亏计算”和“后台盈亏计算”。前者使用理论价值来计算盈亏,其目的是为交易提供参考;后者则使用“逐日盯市”价值,根据市场上实际观测到的价值来计算盈亏。长期来看,这两种计算方法得到的结果是基本相符的,而如果是计算一两天内的盈亏,用这两种计算方法得到的结果有些差异也并不奇怪。

报告盈亏的方式有无数种。你可以计算每个交易台的盈亏,可以计算每个时间段的盈亏,也可以计算每种证券类别的盈亏,还有很多其他计算盈亏的方式。但最基本的计算方法都是一样的。对于期权的盈亏,把盈亏按照各个希腊字母进行分配是一种很有帮助的计

算方式。例如，“delta 盈亏”是一种不错的衡量 delta 对冲效果的指标，“vega 盈亏”则能够衡量波动率管理的水平，对其他希腊值来说也是如此。如果你知道这种将盈亏进行分配的方法，对把你的系统调整得更好会大有帮助。

交互场景

现在，我们已经认识了系统中所有的组件，并且知道它们分别是干什么的，那么让我们来看看它们是如何交互并且完成高频交易这项工作的。如果对每一种可能的交互场景都进行描述，那可就得写好多本书了，所以我们仅仅会举个例子，看看当 E-mini 标普 500 指数（ES）期货合约的价格发生显著变动时会发生些什么。实际上，在发生这种情况时，大量的系统组件都会做出反应，这个场景也会展示两个策略——基于股票相对股指期货变动滞后效应的策略以及股票和期权上的吃掉动作慢的策略，这两个策略在第 3 章中我们都讨论过（在实践中，这个变动还可能会触发指数期货和 ETF 之间的套利策略）。为了简便起见，这里假设我们的策略只在一个远程服务器上运行，并且只涉及很少的几种股票和期权合约。在实践中，这些场景会独立地发生在每一个交易所的远程服务器上，并且会交易许多股票和期权合约。

时刻 1、2、3

在这个场景里，最初发生的事情是 ES 期货合约的价格变动，我们假设它的价格从 1 162.00 美元升到了 1 162.50 美元。图 5-7 描述了最开始各组件为应对这一变化而发生的互动。

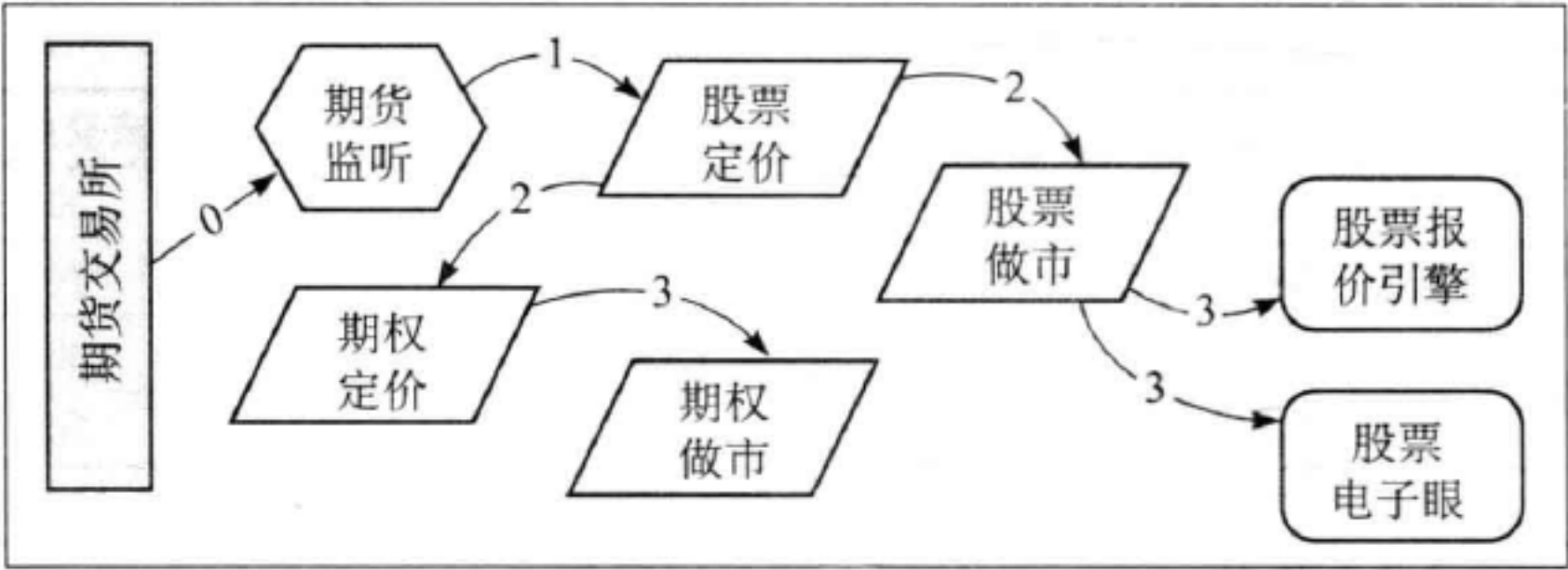


图 5-7 期货价格变动场景：时刻 1 ~ 3

检测 ES 期货合约价格的变动并将其发布给其他系统组件是期货市场监听器的任务，但实际上这项任务并不像它看起来那么简单。期货市场监听器要做的不仅仅是把每一个从交易所传来的新数据发布出来，还需要确认一下新的价格是合理的，而不像是明显的错误数据。除了过滤出那些差的数据之外，它还需要进行一些判断来决定是否将一个数据发布到系统中，因为并不是每个好的数据都需要发布。如果价格在两个价位之上来回快速振动（也称为摆动），就没有必要把所有的这些数据都发布出来，以免下游的组件进行冗余的计算和其他的一些操作。让我们来看一看表 5-4 中所示的虚拟 ES 股指期货合约的最新成交价格流[⊖]（这里的价格既不是申买价也不是申卖价，而是上一笔实际发生交易的成交价格）。

表 5-4 E-mini 标普 500 指数期货价格数据流

时 间	价 格	成交量
下午 2:30:21	1 162.00	1
下午 2:30:21	1 152.00	5
下午 2:30:22	1 162.00	10
下午 2:30:24	1 162.25	300
下午 2:30:24	1 162.00	50
下午 2:30:26	1 162.00	10
下午 2:30:28	1 162.25	50

⊖ 这是一个为了说明而进行了简化的情况。一个真实的时间价格流会包含比这多得多的数据，数据变化频率比这里的秒级数据要高得多。

(续)

时 间	价 格	成交量
下午 2:30:28	1 162.25	100
下午 2:30:28	1 162.50	71
下午 2:30:28	1 162.75	200
下午 2:30:29	1 162.75	50
下午 2:30:29	1 162.75	150
下午 2:30:30	1 162.75	5
下午 2:30:30	1 162.75	12

假设期货市场监听器在开始时刻监听到的 ES 期货合约的最新成交价是 1 162.00 美元，那么对于系统中的其他所有组件来说，在监听器开始评估下一个数据之前，ES 期货合约的当前价就是 1 162.00 美元。下一个时刻，也就是在 2:30:21 时，新来的数据中价格变成了 1 152.00 美元，这和之前一个数据差得实在太远了，因此监听器会认为这是一个差的数据而忽略它。那么对于下一个数据，也就是 2:30:24 时价格为 1 162.25 美元（伴随着一个很大的成交量 350 手）的数据呢？我们的监听器不会发布它，因为紧接着下一个价格回到了 1 162.00。并且再下一个数据还是 1 162.00，这个数据同样会被忽略。但是再看看 2:30:28 时的数据，价格同样向上变动了 0.25 美元，然后在下一个时刻又向上变动了 0.25 美元，在接下来的时刻继续向上变动了 0.25 美元，使价格到达了 1 162.75 美元，然后价格就停在这个数字上了。我们的监听器认为这种变动是真实的，然后它将 1 162.75 美元这个新的 ES 期货合约价格发布给系统的其他组件。这样，系统所获得的新的期货价格就比上一次发布的价格高了 0.75 美元（**注意：**这里为了描述方便故意使用了简化的逻辑。实际的逐笔数据分析和发布的逻辑可能会比这复杂很多）。

让我们再来看看图 5-7，可以看到，股票定价组件在时刻 1 接收到了新的价格。在这种场景下，股票定价组件的任务是根据各股票

在指数中的权重和 β 值（回想一下，我们在之前说过， β 是一种粗略地衡量单只股价对市场整体价位变动敏感性的指标，在这里，我们用指数来代替市场）为标普 500 指数中的每一只成分股计算一个对应的微观价格。在时刻 2，股票定价组件将这些股票新的微观价格发布给系统的其他组件。

接下来，股票做市组件将会使用这些股票新的价格来计算标普 500 指数成分股的买卖报价，而期权定价组件则使用这些新价格在理论价格矩阵中查找期权对应的新的理论价格。在时刻 3，股票做市组件发布新的买卖报价，而期权定价组件发布新的理论价格。值得注意的是，在股票和期权的定价之间会存在一个自然的延迟，因为期权定价是从股票定价上衍生出来的。

时刻 4、5

就像图 5-8 中所描述的，在时刻 4，股票电子眼会执行吃掉动作慢的策略，把市场上那些来不及更新的申买或者申卖委托单吃掉（也就是那些能带来 α 的委托单）。股票报价引擎则同时更新我们的买卖报价，使我们在电子眼吃掉这些委托单后，在相反方向上的报价显得比较有吸引力，从而更容易完成双向交易。同样，期权做市组件也会在时刻 4 更新买卖报价。

在时刻 5，股票交易引擎将会接收到从交易所发送回来的成交反馈。风险管理器会确认每一笔股票的买入交易是否对应了一笔更高价格上的卖出交易（如果不是，或者股票持仓超过了事先设定的阈值，风险管理器就会向电子眼发送消息，让它减少风险暴露）。同样在时刻 5，期权报价引擎会更新它在期权交易所的买卖报价，期权电子眼也会下单吃掉期权市场上的那些来不及对新的市场状况做出反应的交易者的委托单。

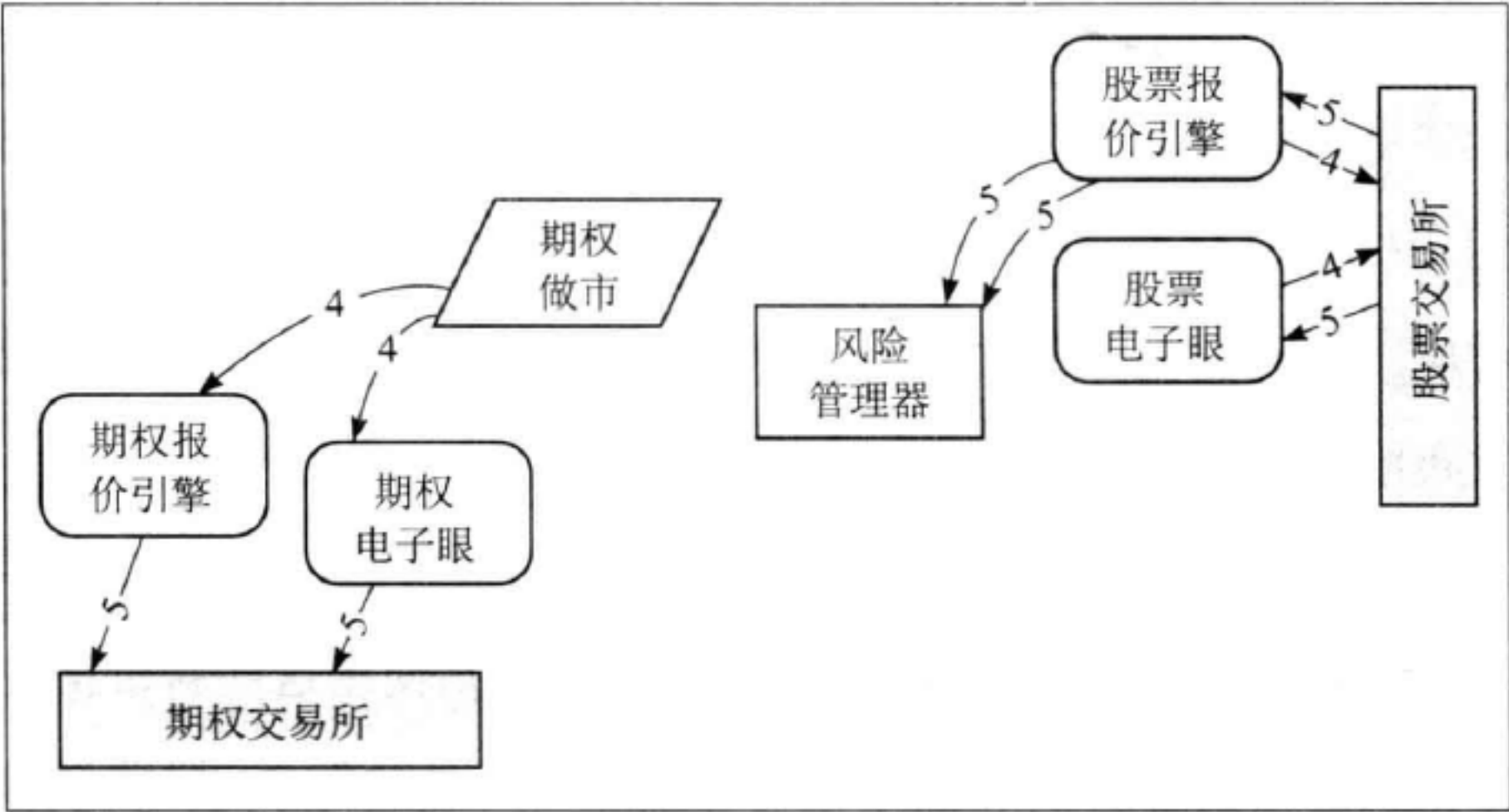


图 5-8 期货价格变动场景：时刻 4、5

时刻 6

如图 5-9 中所描述的，在时刻 6，期权交易引擎将交易完成的信息传递给风险管理和自动对冲器。与股票交易不同，期权交易并不一定需要进行双向的一买一卖的交易（比如，我们没法像股票那样为同一个期权合约找到对应的买方和卖方）。

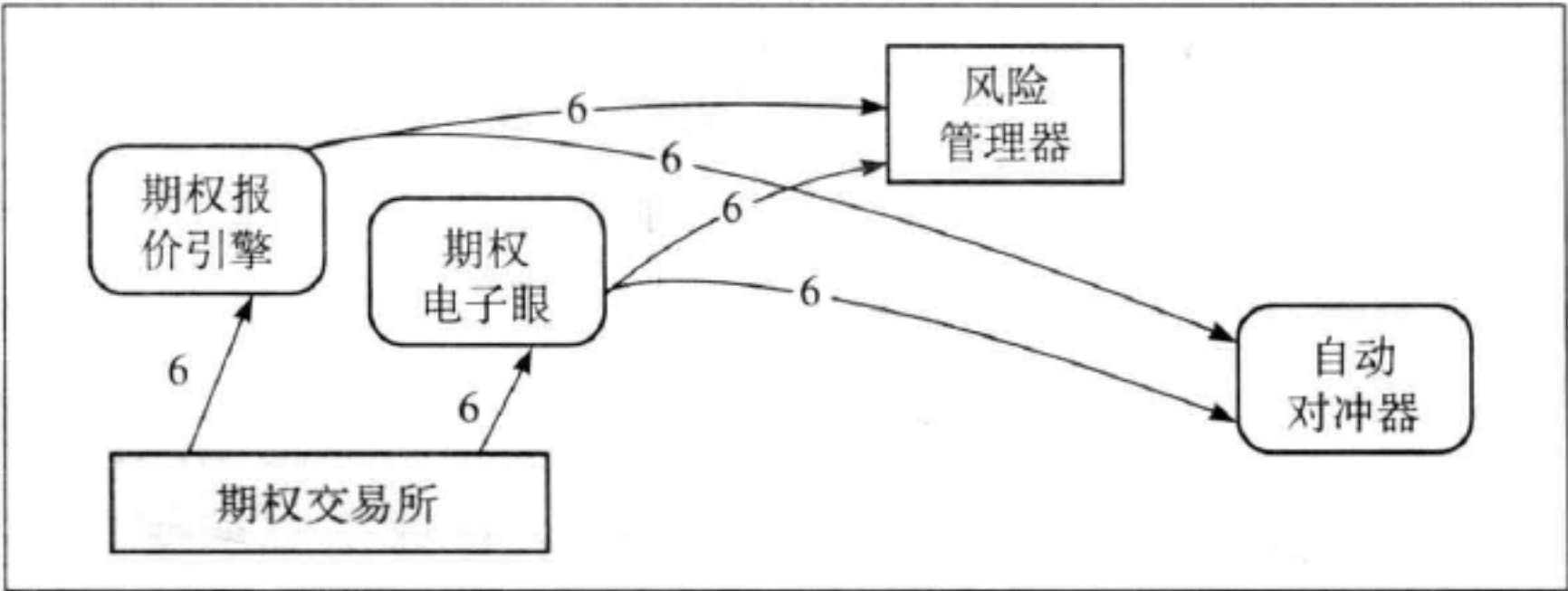


图 5-9 期货价格变动场景：时刻 6

时刻 7、8、9

如图 5-10 所示，在时刻 7，自动对冲器会计算期权交易所带来的 delta 风险暴露，并且会在股票交易所中下单来保持组合的 delta 中

rho：希腊字母之一。度量的是期权价格对无风险利率（通常是 LIBOR）变动的敏感性。

robo trading 机器人交易：见算法交易。

round-trip 双向交易：以一个价格买入，再以另一个更高的价格卖出。做市商和大部分高频交易者每天重复做的工作。另见倒买倒卖。

Russell 2000 罗素 2000：2 000 只股票（最大的 3 000 只股票中最小的 2 000 只）的指数。

S&P500 标普 500：500 只股票的指数，常常作为美国整体股票市场的风向标。

scalp 倒买倒卖[⊖]：以一个价格买入，再以另一个更高的价格卖出。做市商和大部分高频交易者每天重复做的工作。正文也称黄牛策略，另见双向交易。

scratch trade 刮擦交易：以等于证券理论价值的价格执行的交易。另见价值交易。

SEC (Securities and Exchange Commission) 美国证监会：美国负责监管现金资产和期权交易所的监管机构。与 CFTC 的作用类似。

sell-side 卖方：宽泛的指做市商，专家，和高频交易者等，他们为买方提供流动性，并赚取买卖价差和 / 或交易所返还。

server 服务器：一种通用计算机，一般安装在数据中心的机架里，并且通过网络和其他计算机连接在一起。

SFTI (secure financial transaction infrastructure) 安全金融交易基础设施：NYSE 所有的一个低延迟、容错的数据网络，此网络与大部分主要交易所相连。

⊖ Scalp 这个词有人译作“剥头皮”。但剥头皮并不能直观的反映出其所包含的倒买倒卖的意思。因此，我们采用了不同的译法。——译者注

short position 空头持仓：因为净卖出股票，卖出期权，或者卖出期货合约而产生的持仓，证券市场价格上涨时，持仓价值会减少。

short sale 卖空：卖出借入的股票。如果归还股票前股票价格下跌，卖空者就可以获得收益。

size 规模：某份申买或申卖，或交易委托单所对应的股数或者合约数。

size order 大规模委托单：见大额委托单。

specialist 专家：见做市商。

Spider (SPDR)：一只具有很高流动性的跟踪标普 500 指数的 ETF。

spot interest 即期利率：从现在开始的假设性借款的利率。另见远期利率。

spread 价差：通常指申卖价和申买价之间的差。也指为了达到某种收益特征而用多个期权组合起来的期权交易。

SPX：一种流动性非常好的，现金结算的标普 500 股指期权，该期权只在 CBOE 交易。它也是仅有的几只主要由场内交易员进行交易（截止到 2010 年 3 月）的上市期权合约。

statistical arbitrage 统计套利：一种使用统计学和其他数学方法，对交易数据和其他数据进行深度挖掘，以发现稳定的可预测模式的交易技术。

stock 股票：谁都知道股票什么！[⊖]

strategy 策略：期权价差的另一个名字。也指交易者用来获取交易利润的标准技术。

strike price 执行价格：一个事先约定的价格，看涨（看跌）期权的买方有权利而无义务以这个价格买入（卖出）标的证券。

⊖ 好吧，股票是一种美国权益类市场上交易的非常多的金融证券，它代表拥有一个公司的一部分股份。

symbol 代码：股票、期权、或者期货合约的缩写标记。

synthetic 合成的：合成期权指的是用证券组合去模拟一个真实期权的收益，做市商常常用合成期权对冲他们的期权头寸。合成期货指的是用具有相同执行价格和到期时间的看涨期权多头和看跌期权空头（或者是看涨期权空头和看跌期权多头）组成的组合，其收益特征与具有相同标的物的期权合约一致。

take out 取出：成功的打掉申卖或者吃掉申买。

TCP/IP：大部分联网的计算机进行通信时所使用的网络协议。

theoretical value 理论价值：计算出来的给定证券（通常是期权）的价格，在这个价格交易的双方没有哪一方会获利或者亏损。也称为公平市场价值。

theta：希腊字母之一。度量的是期权价格对到期时间变动的敏感性。

throughput 吞吐量：单位时间之内的计算量和 / 或数据传输量。类似于花园水管一段时间之内的出水量。

tick 笔：这个概念有很多种不同的用法：①指交易所发布的价格变化（申买价格和 / 或申买量的变化，申卖价格和 / 或申卖量的变化，最新成交价和 / 或成交量的变化）②指价格水平的变化（例如，股票“涨了几笔”到了新的价格）③指证券价格能够变动的最小增量（分，5分，等等），或者叫最小变动价位。

TOE（TCP/IP offload engine）TCP/IP 减负引擎：一种安装在服务器上的优化硬件设备，它负责把数据传输到网络上（或从网络上传下来），这样处理器就不需要处理这些事情，而可以进行其它工作。

trade-through 越线交易：违反 NBBO 保护规则的交易。例如，一家交易所以高于另一家交易所当前最高申卖价的价格成交了一张买入委托单。

trinomial 三叉（树）：一种允许标的物价格在多重网格（树）上变化的期权定价方法，在树上的每一个分支，价格既可以向上或者向下变动，也可以保持不变。另见二叉树和 FDM。

TV 真实价格：见理论价格。

underlier 标的物：一种决定了其衍生品价值的证券。例如，谷歌股票看涨期权的标的物就是谷歌股票。

value trade 按价值交易：以等于证券理论价值的价格执行的交易。另见刮擦交易。

vega：希腊字母之一。度量的是期权价格对标的物价格波动率变动的敏感性。

volatility 波动率：衡量股票价格变动频率和幅度的指标，通常以标准差的形式表示。另见隐含波动率和历史波动率。

volatility curve 波动率曲线：用来显示波动率水平与一系列执行价格或者到期时间关系的曲线。这种曲线可以画在一张 x-y 坐标图里，其 y 轴或纵轴表示波动率，x 轴表示执行价格或者到期时间。

volatility surface 波动率曲面：把执行价格波动率曲线和到期时间波动率曲线结合在一起的三维数据曲面，其中 x 轴是执行价格，y 轴是到期时间，z 轴表示波动率。

VPN（virtual private network）虚拟专用网络：一组计算机用公用互联网的方式相互连接，但其他电脑无法访问这组计算机。

VWAP（volume-weighted average price）交易量加权平均价：某证券（通常是股票）一段时间内实际交易价格抽样的平均值。

WAN（wide area network）广域网：在地理上大范围分开的计算机组成的网络。

推荐阅读

序号	书号	书名	作者	定价
1	30250	江恩华尔街45年（珍藏版）	（美）威廉 D. 江恩	36.00
2	30248	如何从商品期货贸易中获利（珍藏版）	（美）威廉 D. 江恩	58.00
3	30247	漫步华尔街（原书第9版）（珍藏版）	（美）伯顿 G. 马尔基尔	48.00
4	30244	股市晴雨表（珍藏版）	（美）威廉·彼得·汉密尔顿	38.00
5	30251	以交易为生（珍藏版）	（美）亚历山大·埃尔德	36.00
6	30246	专业投机原理（珍藏版）	（美）维克托·斯波朗迪	68.00
7	30242	与天为敌：风险探索传奇（珍藏版）	（美）彼得 L. 伯恩斯坦	45.00
8	30243	投机与骗局（珍藏版）	（美）马丁 S. 弗里德森	36.00
9	30245	客户的游艇在哪里（珍藏版）	（美）小弗雷德·施韦德	25.00
10	30249	彼得·林奇的成功投资（珍藏版）	（美）彼得·林奇	38.00
11	30252	战胜华尔街（珍藏版）	（美）彼得·林奇	48.00
12	30604	投资新革命（珍藏版）	（美）彼得 L. 伯恩斯坦	36.00
13	30632	投资者的未来（珍藏版）	（美）杰里米 J. 西格尔	42.00
14	30633	超级金钱（珍藏版）	（美）亚当·史密斯	36.00
15	30630	华尔街50年（珍藏版）	（美）亨利·克卢斯	38.00
16	30631	短线交易秘诀（珍藏版）	（美）拉里·威廉斯	38.00
17	30629	股市心理博弈（原书第2版）（珍藏版）	（美）约翰·迈吉	58.00
18	30835	赢得输家的游戏（原书第5版）	（美）查尔斯 D. 埃利斯	36.00
19	30978	恐慌与机会	（美）史蒂芬·韦恩斯	36.00
20	30606	股市趋势技术分析（原书第9版）（珍藏版）	（美）罗伯特 D. 爱德华兹	78.00
21	31016	艾略特波浪理论:市场行为的关键（珍藏版）	（美）小罗伯特 R. 普莱切特	38.00
22	31377	解读华尔街（原书第5版）	（美）杰弗里 B. 利特尔	48.00
23	30635	蜡烛图方法：从入门到精通（珍藏版）	（美）斯蒂芬 W. 比加洛	32.00
24	29194	期权投资策略（原书第4版）	（美）劳伦斯 G. 麦克米伦	128.00
25	30628	通向财务自由之路（珍藏版）	（美）范 K. 撒普	48.00
26	32473	向最伟大的股票作手学习	（美）约翰·波伊克	36.00
27	32872	向格雷厄姆学思考，向巴菲特学投资	（美）劳伦斯 A. 坎宁安	38.00
28	33175	艾略特名著集（珍藏版）	（美）小罗伯特 R. 普莱切特	32.00
29	35212	技术分析（原书第4版）	（美）马丁 J. 普林格	65.00
30	28405	彼得·林奇教你理财	（美）彼得·林奇	36.00

推荐阅读

序号	书号	书名	作者	定价
31	29374	笑傲股市（原书第4版）	（美）威廉·欧奈尔	58.00
32	30024	安东尼·波顿的成功投资	（英）安东尼·波顿	28.00
33	35411	日本蜡烛图技术新解	（美）史蒂夫·尼森	38.00
34	35651	麦克米伦谈期权（珍藏版）	（美）劳伦斯 G. 麦克米伦	80.00
35	35883	股市长线法宝（原书第4版）（珍藏版）	（美）杰里米 J. 西格尔	48.00
36	37812	漫步华尔街（原书第10版）	（美）伯顿 G. 马尔基尔	56.00
37	38436	约翰·聂夫的成功投资（珍藏版）	（美）约翰·聂夫	39.00
38	38520	经典技术分析（上册）	（美）小查尔斯 D. 柯克帕特里克	69.00
39	38519	经典技术分析（下册）	（美）小查尔斯 D. 柯克帕特里克	69.00
40	38433	在股市大崩溃前抛出的人：巴鲁克自传（珍藏版）	（美）伯纳德·巴鲁克	56.00
41	38839	投资思想史	（美）马克·鲁宾斯坦	59.00
42	41880	超级强势股：如何投资小盘价值成长股	（美）肯尼思 L. 费雪	39.00
43	39516	股市获利倍增术（珍藏版）	（美）杰森·凯利	39.00
44	40302	投资交易心理分析	（美）布雷特 N. 斯蒂恩博格	59.00
45	40430	短线交易秘诀（原书第2版）	（美）拉里·威廉斯	49.00
46	41001	有效资产管理	（美）威廉 J. 伯恩斯坦	39.00
47	38073	股票大作手利弗莫尔回忆录	（美）埃德温·勒菲弗	39.80
48	38542	股票大作手利弗莫尔谈如何操盘	（美）杰西 L. 利弗莫尔	25.00
49	41474	逆向投资策略	（美）大卫·德雷曼	59.00
50	42022	外汇交易的10堂必修课	（美）贾里德 F. 马丁内斯	39.00
51	41935	对冲基金奇才：常胜交易员的秘籍	（美）杰克·施瓦格	80.00
52	42615	股票投资的24堂必修课	（美）威廉·欧奈尔	35.00
53	42750	投资在第二个失去的十年	（美）马丁 J. 普林格	49.00
54	44059	期权入门与精通（原书第2版）	（美）爱德华·奥姆斯特德	49.00
55	43956	以交易为生II：卖出的艺术	（美）亚历山大·埃尔德	55.00
56	43501	投资心理学（原书第5版）	（美）约翰 R. 诺夫辛格	49.00
57	44062	马丁·惠特曼的价值投资方法：回归基本面	（美）马丁·惠特曼	49.00
58	44156	巴菲特的投资组合（珍藏版）	（美）罗伯特·哈格斯特朗	35.00
	13303	巴菲特致股东的信	（美）沃伦·巴菲特	38.00
	27497	主动型指数投资	（美）史蒂芬 A. 斯科恩菲尔德	78.00

交易手续费免费办理国内正规商品股指原油期货开户 零佣金 不限资金量和交易量直接申请交易
费用比我们低 10倍赔偿差价 客服QQ/微信：958218088 期货开户和书籍下载请进：<http://www.>

利润以光速奔腾 交易决胜于瞬息

什么是高频交易者？他们是装备技术、预测者的结合体。同时他们也会时不时地运用瞬息即逝的获利机会。他们深谙交易规则，了解大量套利策略，运用高科技设备，以微秒为单位进行交易。积小利，而终至万金。

这个群体正日益壮大，而本书则试图揭开这些神秘宽客的面纱，将这台赚钱机器的轮廓展示给广大读者。

最具实战经验的作者，最具专业知识的译者，展现出来的高频交易也最原汁原味。对于高频交易的研究者、管理者以及监管者，我都高度推荐此书。

——罗旭峰 南华期货股份有限公司总经理

McGraw-Hill
全球智慧中文化
<http://www.mheducation.com>



客服热线：(010) 68995261 88361066
购书热线：(010) 68326294 88379649 68995259
投稿热线：(010) 88379007

数字阅读：www.hzmedia.com.cn
华章网站：www.hzbook.com
网上购书：www.china-pub.com

上架指导：投资/证券

ISBN 978-7-111-44658-3

9 787111 446583 >

定价：40.00元